

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 43.3333 Seconds  
(without alignments)  
30.750 Million cell updates/sec

Title: US-09-673-785D-8

Perfect score: 54

Sequence: 1 CVIGXSGDRC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

```

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	10	AAV52143	Mouse EGF derived
2	54	100.0	47	AG66047	Mouse EGF motif se
3	54	100.0	53	AA40315	Epidermal growth f
4	54	100.0	53	AA40315	Amino acid sequenc
5	54	100.0	53	AA40315	Modified murine epi
6	54	100.0	53	AA40315	Synthetic human/mo
7	54	100.0	53	AA40315	Chimeric epidermal
8	54	100.0	53	AA40315	Chimeric epidermal
9	54	100.0	53	AA40315	Mouse epidermal gr
10	54	100.0	53	AA40315	Epidermal growth f

11	54	100.0	53	19	AAW50140	Epidermal growth f
12	54	100.0	53	22	AAW50140	Human EGF. Homo s
13	54	100.0	53	22	AAW50140	Epidermal growth f
14	54	100.0	54	23	AAU76704	Mouse epidermal gr
15	54	100.0	54	23	AAU76704	Modified mouse epi
16	54	100.0	56	7	AAU76704	Epidermal growth f
17	54	100.0	117	19	AAW68455	Recombinant human
18	54	100.0	376	22	AAU02933	Angiotensin conver
19	54	100.0	376	22	AAU02933	Angiotensin conver
20	54	100.0	444	22	AAU02933	Angiotensin conver
21	54	100.0	576	22	AAU02933	Angiotensin conver
22	54	100.0	576	22	AAU02933	Angiotensin conver
23	54	100.0	1073	22	AAU02933	Angiotensin conver
24	49	90.7	2146	22	ABG62317	Drosophila melanog
25	47	87.0	1080	6	AAU50296	Mouse epidermal gr
26	45	83.3	14	18	AAW34460	Chimeric epidermal
27	44	81.5	46	23	AAW66045	Mouse NRG-2 EGF-li
28	44	81.5	48	15	AAU57106	N terminal of rat
29	44	81.5	48	20	AAU01790	Rat epidermal grow
30	44	81.5	71	9	AAU81372	Rat epidermal grow
31	44	81.5	73	9	AAU81367	Rat epidermal grow
32	44	81.5	298	23	AAU11636	Human Neuregulin-2
33	44	81.5	308	22	AAW35196	Human transmembran
34	44	81.5	366	22	AAW35196	Human transmembran
35	44	81.5	366	22	AAW35202	Human transmembran
36	44	81.5	754	18	AAW27536	Rat cerebellum der
37	44	81.5	1296	23	ABG66702	Human novel polype
38	44	81.5	1296	23	ABG66756	Human novel polype
39	43	79.6	17	13	AAU21705	EGF fragment. Syn
40	43	79.6	17	13	AAU21705	Substituted EGF fr
41	43	79.6	17	18	AAW11619	Target peptide fro
42	43	79.6	17	18	AAW11620	Target peptide fro
43	43	79.6	22	18	AAW11621	Target peptide fro
44	43	79.6	22	18	AAW11622	Target peptide fro
45	43	79.6	35	22	AAW98651	Epidermal Growth F

## ALIGNMENTS

RESULT 1  
AAV52143  
ID AAV52143 standard; peptide: 10 AA.  
AC AAV52143;  
XX  
DT 28-JAN-2000 (first entry)  
XX Mouse EGF derived peptide for targetting laminin receptor.

DE  
KW Epidermal growth factor; EGF; laminin receptor; angiogenesis;  
KW medicament; wound healing; retinopathy of immaturity; metastatic cancer;  
KW candida infection; leishmania; trichomonas vaginalis.

XX Mus sp.  
XX OS  
XX  
XX Key Location/Qualifiers  
XX Modified-site 1 /note= "Acetyl-Cys (S-ACM)"  
XX Modified-site 10 /note= "Cys (S-ACM)-NH2"  
XX  
XX W0954356-A1.  
XX  
XX 28-OCT-1999.  
XX  
XX 21-APR-1999; 99WO-GB01211.  
XX  
XX 22-APR-1998; 98GB-0008407.  
XX (UYBE-) UNIV QUEENS BELFAST.  
XX Nelson J, Walker B, McFerran N, Harriott P;  
PI

XX WPI; 2000-013229/01.  
XX New peptide derived from murine epidermal growth factor (mEGF) -  
XX Claim 4; Page 28; 35pp; English.  
XX This is a peptide derived from mouse epidermal growth factor (EGF)  
XX residues 33-42. This peptide is used in the invention to prepare a  
XX composition to target laminin receptors. EGF derived peptides inhibit  
XX blood vessel formation through their antagonism of the high affinity 67kd  
XX laminin receptor found on endothelial cells. The peptide is modified  
XX from the natural sequence to prevent protease attack. The peptide is used  
XX in the preparation of a medicament for binding to laminin receptors as an  
XX (ant)agonist. The medicament is also useful for healing endothelial cell  
XX wounds and treating angiogenic diseases, especially retinopathy of  
XX immaturity. Other diseases treated include metastatic cancer,  
XX Candida spp. infection, and parasitic infestations like leishmania and  
XX trichomonas vaginalis. The peptide are anti-angiogenic in human models.  
XX The peptides also inhibit both laminin- and EGF-stimulated angiogenesis,  
XX and prevent tumour cell attachment to basement membranes.  
XX Sequence 10 AA;

Query Match 100.0%; Score 54; DB 21; Length 10;  
Best Local Similarity 90.0%; Pred. NO. 0.074; 0; Indels 0; Gaps 0;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CVIGXSGDRC 10  
DB 1 CVIGYSGDRC 10  
||||:|||||

RESULT 2  
AAG66047  
ID. AAG66047 standard; peptide; 47 AA.  
AC AAG66047;  
XX 27-FEB-2002 (first entry)  
XX Mouse EGF motif sequence.  
XX ErbB-4; neuregulin-4; NRG-4; pro-NRG-4; neuroprotective; vulnary;  
XX cerebroprotective; vasotropic; antiparkinsonian; anticonvulsant;  
XX cytostatic; nootropic; EGF; NRG-2.  
XX Mus musculus.  
XX WO200181540-A2.  
XX 01-NOV-2001.  
XX 20-APR-2001; 2001WO-IL00371.  
XX 21-APR-2000; 2000US-0553769.  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX Harari D, Yarden Y;  
XX WPI; 2002-041398/05.  
XX Novel ErbB-4 ligand, referred as neuregulin (NRG)-4 and polynucleotide  
XX sequences encoding NRG-4, useful for upregulating or downregulating  
XX ErbB-4 receptor activity to treat Alzheimer's disease, stroke, gastric  
XX cancer -  
XX Disclosure; Fig 1c; 153pp; English.  
XX The invention relates to a novel ErbB-4 ligand, neuregulin-4 (NRG-4).  
XX NRG-4 binds to mammalian ErbB-4 receptor and can be expressed by standard  
XX recombinant methodology. Pharmaceutical compositions comprising NRG-4 are

CC useful for regulating an endogenous protein affecting ErbB-4 receptor  
CC activity in vivo. They are also useful for treating or preventing a  
CC disease condition or syndrome associated with dysregulation of an  
CC endogenous protein affecting ErbB-4 receptor activity, e.g., amyotrophic  
CC lateral sclerosis (Lou Gehrig's disease), Bell's palsy, spinal muscular  
CC atrophy, brain trauma, stroke, ischemia, Alzheimer's disease, Parkinson's  
CC disease, epilepsy, multiple sclerosis, Huntington's chorea, Down's  
CC syndrome, nerve deafness, neuropathy, muscular dystrophy, extramammary  
CC Paget's disease, gastric, pancreatic, prostate, breast and ovarian  
CC cancer, cervical carcinoma, endometrial adenocarcinoma, pancreatic D  
CC cells-somatostatinoma and Zollinger-Ellison syndrome. The agent comprised  
CC in the pharmaceutical composition includes a polypeptide (e.g., a soluble  
CC ligand binding domain of ErbB-4 i.e., IgB4; or a monoclonal, polyclonal,  
CC humanized, single chain antibody or an immunoreactive derivative of an  
CC antibody) capable of binding the endogenous protein affecting ErbB-4  
CC receptor activity. Traceable synthetic/recombinant NRG-4-tagged molecules  
CC can serve as a diagnostic tool in which cells binding NRG-4 can be  
CC measured. Sequences AAG66044-53 represent the EGF-like motifs of various  
CC growth factors.  
XX Sequence 47 AA;

Query Match 100.0%; Score 54; DB 23; Length 47;  
Best Local Similarity 90.0%; Pred. NO. 0.29;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CVIGXSGDRC 10  
DB 33 CVIGYSGDRC 42  
||||:|||||

RESULT 3  
AAP40315  
ID. AAP40315 standard; protein; 53 AA.  
XX AAP40315;  
XX 22-JUL-1992 (first entry)  
XX Epidermal growth factor tripentacontapeptide.  
XX EGF.  
XX Synthetic.  
XX Key Location/Qualifiers  
XX Disulfide-bond 6...20  
XX Disulfide-bond 14...31  
XX Disulfide-bond 33...42  
XX JF59027858-A.  
XX 14-FEB-1984.  
XX 05-AUG-1982; 82JP-0137128.  
XX 05-AUG-1982; 82JP-0137128.  
XX (NNSH ) NIPPON SHINYAKU KK.  
XX WPI; 1984-072465/12.  
XX Synthesis of epidermal growth factor polypeptide - by condensation  
XX of protected smaller peptide sequences, deprotection then oxidn. to  
XX cyclise.  
XX Claim1; Page 1; 8pp; Japanese.  
XX The amino acid sequence is that of an epidermal growth factor  
XX tripentacontapeptide which is synthesised by condensation of  
XX protected smaller peptide sequences. This method produces the  
XX peptide smoothly, with high purity and yield.

SQ Sequence 53 AA;

Query Match 100.0%; Score 54; DB 5; Length 53;  
 Best Local Similarity 90.0%; Pred. No. 0.33;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
 ||||:||||  
 Db 33 CVIGYSGDRC 42

## RESULT 4

AAP91658  
 ID AAP91658 standard; protein; 53 AA.

XX  
 AC AAP91658;

XX  
 DT 29-JUN-1990 (first entry)

XX  
 DE Amino acid sequence for naturally occurring epidermal growth factor (EGF).

XX  
 KW Epidermal growth factor; angiogenesis; synthetic peptide.

XX  
 PN W08901489-A.

XX  
 PD 23-FEB-1989.

XX  
 PF 10-AUG-1988; 88WO-AU00300.

XX  
 PR 10-AUG-1987; 87AU-0003629.

XX  
 PA (CSIR ) COMMONWEALTH SCIENT ORG.

XX  
 PI McAuslan BR;

XX  
 DR WPI; 1989-068852/09.

XX  
 PT Synthetic peptide active in stimulating angiogenesis -  
 PT has sequences corresponding to amino acid sequences occurring in  
 PT epidermal growth factor.

XX  
 PS Fig 1; 1/1; 11pp; English.

XX  
 CC The inventors claim synthetic peptides which correspond to sequences  
 CC occurring in EGF, but excluding EGF. The synthetic peptides  
 CC correspond to active sequences from EGF fragments: 3-14, 12-14, 12-15,  
 CC 3-10, 29-37 and 33-37 (Claims 2-7 and AAP91659-P91664 resp.). The  
 CC peptides are angiogenic. Their relative shortness means that they pose  
 CC fewer synthesis problems than the entire EGF molecule.

SQ Sequence 53 AA;

Query Match 100.0%; Score 54; DB 10; Length 53;  
 Best Local Similarity 90.0%; Pred. No. 0.33;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
 ||||:||||  
 Db 33 CVIGYSGDRC 42

## RESULT 5

AAR08007  
 ID AAR08007 standard; protein; 53 AA.

XX  
 AC AAR08007;

XX  
 DT 25-FEB-1991 (first entry)

XX  
 DE Modified murine epidermal growth factor.

XX  
 KW Modified murine epidermal growth factor; stability; storage;

KW epithelial wounds; gastric acid secretion.

XX Key Location/Qualifiers

FT Misc-difference 11.11  
 FT /label= E, N, Q, A, K

XX  
 PN W09013570-A.

XX  
 PD 15-NOV-1990.

XX  
 PF 09-MAY-1990; 90WO-US02600.

XX  
 PR 12-MAY-1989; 89US-0351773.

XX  
 PA (CHIR-) CHIRON CORP.

XX  
 PI Nascimento CG, Medina-Selby A;

XX  
 DR WPI; 1990-361427/48.

XX  
 PT Human epidermal growth factor - is substituted at position 11 for  
 PT greater stability and improved storage life.

XX  
 PS Claim 9; Page 25; 32pp; English.

XX  
 CC The human rEGF is used to treat oversecretion of gastric acid or an  
 CC epithelial wound. EGF is modified to increase its chemical  
 CC stability. Its storage life is improved without diminishing its  
 CC biological activity. The proteins may be prepared by traditional  
 CC chemical or recombinant means.  
 CC See also AAR08004.

XX  
 SQ Sequence 53 AA;

Query Match 100.0%; Score 54; DB 11; Length 53;  
 Best Local Similarity 90.0%; Pred. No. 0.33;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
 ||||:||||  
 Db 33 CVIGYSGDRC 42

## RESULT 6

AAR67275  
 ID AAR67275 standard; peptide; 53 AA.

XX  
 AC AAR67275;

XX  
 DT 01-AUG-1995 (first entry)

XX  
 DE Synthetic human/mouse EGF.

XX  
 KW Cell growth factor; viscoelastic solution; fibroblastic growth factor;  
 KW EGF; epidermal growth factor; EGF; buffered solution; lubrication;  
 KW carboxypropyl methylcellulose; cellulose gum; dextran; dextran sulphate;  
 KW chondroitin sulphate; sodium hyaluronate; osmolality; mitogenic;  
 KW wound healing; cell protection; cell coating; surgery; tissue space;  
 KW hydroxypropyl methylcellulose; manipulation.

XX  
 OS Synthetic.

XX  
 FH Key Location/Qualifiers

FT Misc-difference 37  
 FT /note= "Given in the specification as Try"

XX  
 PN US5366964-A.

XX  
 PD 22-NOV-1994.

XX  
 PF 15-DEC-1988; 88US-0284533.

XX  
 PR 15-DEC-1988; 88US-0284533.





PT Chimeric epidermal growth factor proteins - and DNA molecules for  
 PT their recombinant production  
 XX  
 PS Claim 6; Page 26; 40pp; English.  
 XX This is a chimeric epidermal growth factor (EGF) protein ABC. This  
 CC chimeric EGF protein contains sequences derived from human and mouse  
 CC EGF sequences. The chimeric EGF proteins can be recombinantly produced  
 CC by culturing a microorganism of the genus *Bacillus* transformed by an  
 CC expression vector containing the encoding DNA molecules. The recombinant  
 CC chimeric EGF proteins exhibit a wide variety of physiological activities  
 CC similar to those of naturally occurring EGF.  
 XX  
 SQ Sequence 53 AA;  
 Query Match 100.0%; Score 54; DB 18; Length 53;  
 Best Local Similarity 90.0%; Pred. No. 0.33;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDRC 10  
 Db 33 CVIGYSGDRC 42  
 RESULT 9  
 AAW50134  
 ID AAW50134 standard; protein; 53 AA.  
 XX  
 AC AAW50134;  
 XX  
 DT 08-JUL-1998 (first entry)  
 XX  
 DE Mouse epidermal growth factor.  
 XX  
 KW Mouse; epidermal growth factor; EGF; isolation; recombinant;  
 KW ion exchange chromatography; *Bacillus* brevis.  
 XX  
 OS Mus sp.  
 XX  
 PN AU9728698-A.  
 XX  
 PD 05-FEB-1998.  
 XX  
 PF 17-JUL-1997; 97AU-0028698.  
 XX  
 PR 01-AUG-1996; 96JP-0218109.  
 XX  
 PA (HGET ) HIGETA SHOYU KK.  
 XX  
 PI Miyauchi A, Nemoto A, Warren B;  
 XX  
 DR WPI; 1998-131057/13.  
 XX  
 PT Isolation of recombinant epidermal growth factor from whole broth -  
 PT by fluidised-bed ion-exchange chromatography  
 XX  
 PS Claim 6; Page 26; 38pp; English.  
 XX  
 CC The present sequence represents mouse epidermal growth factor (EGF)  
 CC from the present invention. The present invention describes a method for  
 CC the isolation of recombinant epidermal growth factor (EGF) from whole  
 CC broth. The method comprises: (a) passing a culture containing the  
 CC protein upwards through a column containing a fluidised bed of ion  
 CC exchanger, washing the ion exchanger to remove non-adsorbed material,  
 CC allowing the ion exchanger to settle, and eluting adsorbed protein by  
 CC passing an eluant downwards through the column; (b) filtering the eluate  
 CC through a membrane with a molecular weight cutoff of 10000, and (c)  
 CC concentrating the filtrate on a membrane with a molecular weight cutoff  
 CC of 5000. The method is for isolating recombinant human, mouse, pig or  
 CC rat EGF, or an EGF analogue with one of four defined amino acid sequences  
 CC (given in the specification), from *Bacillus* brevis cultures. The EGF  
 CC proteins can be recovered in high yield and high purity without the need  
 CC to pretreat the culture broth to remove the cells, e.g. by costly

CC centrifugation or membrane filtration.  
 XX  
 SQ Sequence 53 AA;  
 Query Match 100.0%; Score 54; DB 19; Length 53;  
 Best Local Similarity 90.0%; Pred. No. 0.33;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDRC 10  
 Db 33 CVIGYSGDRC 42  
 RESULT 10  
 AAW50139  
 ID AAW50139 standard; protein; 53 AA.  
 XX  
 AC AAW50139;  
 XX  
 DT 08-JUL-1998 (first entry)  
 XX  
 DE Epidermal growth factor analogue 3.  
 XX  
 KW Epidermal growth factor; EGF; isolation; recombinant;  
 KW ion exchange chromatography; *Bacillus* brevis.  
 XX  
 OS Unidentified.  
 XX  
 PN AU9728698-A.  
 XX  
 PD 05-FEB-1998.  
 XX  
 PF 17-JUL-1997; 97AU-0028698.  
 XX  
 PR 01-AUG-1996; 96JP-0218109.  
 XX  
 PA (HGET ) HIGETA SHOYU KK.  
 XX  
 PI Miyauchi A, Nemoto A, Warren B;  
 XX  
 DR WPI; 1998-131057/13.  
 XX  
 PT Isolation of recombinant epidermal growth factor from whole broth -  
 PT by fluidised-bed ion-exchange chromatography  
 XX  
 PS Claim 6; Page 28; 38pp; English.  
 XX  
 CC The present sequence represents a protein which has epidermal growth  
 CC factor (EGF) like activity, from the present invention. The present  
 CC invention describes a method for the isolation of recombinant epidermal  
 CC growth factor (EGF) from whole broth. The method comprises: (a) passing  
 CC a culture containing the protein upwards through a column containing a  
 CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-  
 CC adsorbed material, allowing the ion exchanger to settle, and eluting  
 CC adsorbed protein by passing an eluant downwards through the column; (b)  
 CC filtering the eluate through a membrane with a molecular weight cutoff  
 CC of 10000, and (c) concentrating the filtrate on a membrane with a  
 CC molecular weight cutoff of 5000. The method is for isolating recombinant  
 CC human, mouse, pig or rat EGF, or an EGF analogue with one of four  
 CC defined amino acid sequences (given in the specification), from *Bacillus*  
 CC brevis cultures. The EGF proteins can be recovered in high yield and  
 CC high purity without the need to pretreat the culture broth to remove the  
 CC cells, e.g. by costly centrifugation or membrane filtration.  
 XX  
 SQ Sequence 53 AA;  
 Query Match 100.0%; Score 54; DB 19; Length 53;  
 Best Local Similarity 90.0%; Pred. No. 0.33;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDRC 10  
 Db 33 CVIGYSGDRC 42

```

RESULT 11
AAW50140
ID AAW50140 standard; protein; 53 AA.
XX AC AAW50140;
XX DT 08-JUL-1998 (first entry)
XX DE Epidermal growth factor analogue 4.
XX KW Epidermal growth factor; EGF; isolation; recombinant;
XX KW ion exchange chromatography; Bacillus brevis.
XX OS Unidentified.
XX PN AU9728698-A.
XX PD 05-FEB-1998.
XX PF 17-JUL-1997; 97AU-0028698.
XX PR 01-AUG-1996; 96JP-0218109.
XX PA (HGET ) HIGETA SHOYU KK.
XX PI Miyauchi A, Nemoto A, Warren B;
XX DR WPI; 1998-131057/13.
XX PT Isolation of recombinant epidermal growth factor from whole broth -
XX PT by fluidised-bed ion-exchange chromatography
XX PS Claim 6; Page 29; 38pp; English.
XX CC The present sequence represents a protein which has epidermal growth
XX CC factor (EGF) like activity, from the present invention. The present
XX CC invention describes a method for the isolation of recombinant epidermal
XX CC growth factor (EGF) from whole broth. The method comprises: (a) passing
XX CC a culture containing the protein upwards through a column containing a
XX CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-
XX CC adsorbed material, allowing the ion exchanger to settle, and eluting
XX CC adsorbed protein by passing an eluant downwards through the column; (b)
XX CC filtering the eluate through a membrane with a molecular weight cutoff
XX CC of 10000, and (c) concentrating the filtrate on a membrane with a
XX CC molecular weight cutoff of 5000. The method is for isolating recombinant
XX CC human, mouse, pig or rat EGF, or an EGF analogue, with one of four
XX CC defined amino acid sequences (given in the specification), from Bacillus
XX CC brevis cultures. The EGF proteins can be recovered in high yield and
XX CC high purity without the need to pretreat the culture broth to remove the
XX CC cells, e.g. by costly centrifugation or membrane filtration.
XX SQ Sequence 53 AA;

Query Match 100.0%; Score 54; DB 19; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.33;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db |||||
33 CVIGXSGDRC 42

RESULT 12
AAB37612
ID AAB37612 standard; protein; 53 AA.
XX AC AAB37612;
XX DT 27-FEB-2001 (first entry)
XX DE Human EGF.

XX KW Merozoite surface protein; protozoacide; vaccine; malaria; human; EGF;
XX KW epidermal growth factor.
XX OS Homo sapiens.
XX PN WO200063245-A2.
XX PD 26-OCT-2000.
XX PF 20-APR-2000; 2000WO-GB01558.
XX PR 20-APR-1999; 99GB-0009072.
XX PR 13-MAY-1999; 99US-0311817.
XX PR 25-MAY-1999; 99CA-2271451.
XX PA (MED1-) MEDICAL RES COUNCIL.
XX PI Holder A, Birdsell B, Feeney J, Morgan W, Syed S, Uthaipibull C;
XX DR WPI; 2001-015762/02.
XX PF Novel variants of the C-terminal fragment of Plasmodium merozoite
XX PT surface protein-1, useful as vaccines for treating or preventing
XX PT malaria -
XX XX Disclosure; Fig 1; 126pp; English.
XX CC The present invention relates to non-natural variants of a C-terminal
XX CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
XX CC non-natural variants have reduced affinity for at least 1 antibody
XX CC capable of blocking a second antibody that inhibits the proteolytic
XX CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
XX CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
XX CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of
XX CC the present invention are useful for immunising a mammal against malaria,
XX CC and can be used to treat malaria. The present sequence is human
XX CC epidermal growth factor (EGF). This sequence was used in a sequence
XX CC homology comparison with the wild-type MSP-1 protein from P. falciparum
XX CC (see AAB37608), which was used to generate the variants of the present
XX CC invention.
XX SQ Sequence 53 AA;

Query Match 100.0%; Score 54; DB 22; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.33;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db |||||
33 CVIGXSGDRC 42

RESULT 13
AAE15714
ID AAE15714 standard; Protein; 53 AA.
XX AC AAE15714;
XX DT 12-MAR-2002 (first entry)
XX DE Epidermal growth factor (EGF) used in exemplification of the invention.
XX KW Humanised form; monoclonal antibody alpha 340; gene therapy;
XX KW epidermal growth factor receptor; EGF; cancer; colorectal; lung; breast;
XX KW gastric; ovarian; immune response; cytostatic; cell growth; apoptosis;
XX KW inhibitor.
XX OS Unidentified.
XX PN WO200188138-A1.
XX PD 22-NOV-2001.

```

XX PF 21-MAY-2001; 2001WO-GB02226.  
 XX PR 19-MAY-2000; 2000GB-0011981.  
 XX PR 24-AUG-2000; 2000GB-0020794.  
 XX PA (SCAN-) SCANCELL LTD.  
 XX PI Ellis JRM, Durrant LG;  
 XX PI WPI; 2002-062384/08.  
 XX DR  
 XX PT New humanized form of mouse monoclonal antibody 340 which binds to  
 XX PT epidermal growth factor receptor and inhibits binding of growth factor,  
 XX PT useful for treating colorectal, lung, breast, gastric and ovarian  
 XX PT cancer -  
 XX PS Example 5; Fig 12; 53pp; English.  
 XX CC The present invention relates to a humanised form of the antibody 340 (a  
 CC mouse monoclonal antibody which binds to epidermal growth factor (EGF)  
 CC receptor and inhibits binding of EGF), obtainable from the cell line  
 CC deposited with the ECACC under accession number 97021428. The humanised  
 CC form of the antibody 340 is useful in gene therapy, medicine and in the  
 CC manufacture of a medicament for treatment or prophylaxis of cancer. The  
 CC invention is useful for treating colorectal, lung, breast, gastric or  
 CC ovarian cancers or also for preventing the recurrence of cancer after  
 CC initial treatment or surgery. The invention is also useful for enhancing  
 CC a protective immune response against cancer by optimised immunisation  
 CC schedules. The humanised form of the antibody 340 has reduced  
 CC immunogenicity but shows similar binding to cells expressing EGF  
 CC receptor, as the original murine antibody and has increased ability to  
 CC inhibit the growth of EGF receptor expressing cells. The invention is  
 CC used as cell growth and apoptosis inhibitor. The present sequence  
 CC is epidermal growth factor (EGF) which is used in the exemplification of  
 CC the invention.  
 XX SQ Sequence 53 AA;  
 Query Match 100.0%; Score 54; DB 23; Length 53;  
 Best Local Similarity 90.0%; Pred. No. 0.33;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDRC 10  
 ||||:||||  
 Db 33 CVIGYSGDRC 42  
 RESULT 14  
 AAU76704  
 ID AAU76704 standard; Protein; 54 AA.  
 AC AAU76704;  
 XX  
 XX 21-MAY-2002 (first entry)  
 XX Mouse epidermal growth factor (EGF).  
 XX DE  
 XX KW Mouse; epidermal growth factor; receptor; EGF;  
 XX KW chimeric polynucleotide; directed evolution; chimeraogenesis;  
 XX KW recombination.  
 XX OS Mus sp.  
 XX OS  
 XX FH Key Location/Qualifiers  
 XX FT Misc-difference 54 /label= Unknown  
 XX FT /note= "Encoded by TAA"  
 XX PN WO200206469-A2.  
 XX XX 24-JAN-2002.  
 XX XX

PF 18-JUL-2001; 2001WO-US22640.  
 XX PR 18-JUL-2000; 2000US-218921P.  
 XX PR 18-JUL-2000; 2000US-219085P.  
 XX PR 19-OCT-2000; 2000US-0691873.  
 XX PR 19-OCT-2000; 2000US-0692732.  
 XX PA (ENCH-) ENCHIRA BIOTECHNOLOGY CORP.  
 XX PI Coco WM, Ensell LP, Arensdorf JJ;  
 XX PI WPI; 2002-179795/23.  
 XX DR N-PSDB; ABK10313.  
 XX PT Forming chimeric polynucleotide by contacting single-stranded (ss)  
 XX PT scaffold fragment, and donor fragment populations to form hybridised  
 XX PT complex having ss regions which are filled-in, and ligating adjacent  
 XX PT fragments -  
 XX PS Example 2; Fig 3; 65pp; English.  
 XX CC The invention describes a method of forming (M1) chimeric  
 CC polynucleotides (I). The method comprises contacting a population of  
 CC single-stranded scaffold fragments (SF) with population of donor  
 CC fragments (DF) to form a complex (II) comprising at least one SF  
 CC hybridised to two DFs. (II) is treated so that single stranded regions of  
 CC (II) are filled-in and adjacent fragments are ligated. (M1) is useful for  
 CC a directed evolution process which involves forming a library of (I)  
 CC that can be screened for a characteristic of interest. Subsequent rounds  
 CC of directed evolution can produce chimeric polynucleotides with an  
 CC improved characteristic of interest. The methods facilitate the  
 CC generation of chimeric polynucleotides and do not require hybridising  
 CC donor fragments to a target- or full-length template. Because the  
 CC chimeraogenesis process does not rely upon a contiguous, full-length  
 CC template, it is unnecessary to modify a template to facilitate its  
 CC removal. This is the amino acid sequence of the mouse epidermal growth  
 CC factor (EGF) used to demonstrate a method of in vitro recombination  
 CC described in the method of the invention.  
 XX SQ Sequence 54 AA;  
 Query Match 100.0%; Score 54; DB 23; Length 54;  
 Best Local Similarity 90.0%; Pred. No. 0.33;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDRC 10  
 ||||:||||  
 Db 33 CVIGYSGDRC 42  
 RESULT 15  
 AAU76706  
 ID AAU76706 standard; Protein; 54 AA.  
 AC AAU76706;  
 XX  
 XX 21-MAY-2002 (first entry)  
 XX DE Modified mouse epidermal growth factor (EGF).  
 XX KW Mouse; epidermal growth factor; receptor; EGF;  
 XX KW chimeric polynucleotide; directed evolution; chimeraogenesis;  
 XX KW recombination.  
 XX OS Mus sp.  
 XX OS Synthetic.  
 XX FH Key Location/Qualifiers  
 XX FT Misc-difference 38 /note= "Encoded by ACC"  
 XX PN WO200206469-A2.  
 XX XX

PD 24-JAN-2002.  
XX  
PF 18-JUL-2001; 2001WO-US22640.  
XX  
XX 18-JUL-2000; 2000US-218921P.  
PR 18-JUL-2000; 2000US-219085P.  
PR 19-OCT-2000; 2000US-0691873.  
PR 19-OCT-2000; 2000US-0692732.  
XX  
PA (ENCH-) ENCHIRA BIOTECHNOLOGY CORP.  
XX  
PI Coco WM, Enceill LP, Arensdorf JJ;  
XX  
DR WPI; 2002-179795/23.  
DR N-PSDB; ABK10315.  
XX  
PT Forming chimeric polynucleotide by contacting single-stranded (ss)  
PT scaffold fragment, and donor fragment populations to form hybridised  
PT complex having ss regions which are filled-in, and ligating adjacent  
PT fragments -  
XX  
PS Example 2; Fig 3; 65pp; English.  
XX  
CC The invention describes a method of forming (M1) chimeric  
CC polynucleotides (I). The method comprises contacting a population of  
CC single-stranded scaffold fragments (SF) with population of donor  
CC fragments (DF) to form a complex (II) comprising at least one SF  
CC hybridised to two Dfs. (II) is treated so that single stranded regions of  
CC (II) are filled-in and adjacent fragments are ligated. (M1) is useful for  
CC a directed evolution process which involves forming a library of (I)  
CC that can be screened for a characteristic of interest. Subsequent rounds  
CC of directed evolution can produce chimeric polynucleotides with an  
CC improved characteristic of interest. The methods facilitate the  
CC generation of chimeric polynucleotides and do not require hybridising  
CC donor fragments to a target- or full-length template. Because the  
CC chimeragenesis process does not rely upon a contiguous, full-length  
CC template, it is unnecessary to modify a template to facilitate its  
CC removal. This is the amino acid sequence of a mouse epidermal growth  
CC factor (EGF) modified to make it as similar as possible to the human  
CC protein (AAU76705) and used to demonstrate a method of in vitro  
CC recombination described in the method of the invention.  
XX  
SQ Sequence 54 AA;  
Query Match 100.0%; Score 54; DB 23; Length 54;  
Best Local Similarity 90.0%; Pred. No. 0.33;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CVICXSGDRC 10  
| | | | |  
Db 33 CVIGYSGDRC 42

Search completed: July 2, 2003, 07:30:54  
Job time : 44.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 10.6667 Seconds  
(without alignments)  
27.584 Million cell updates/sec

Title: US-09-673-785D-8  
Perfect score: 54  
Sequence: 1 CVIGXSGDRC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCRUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	12	1	US-08-597-545-7
2	54	100.0	12	1	US-08-457-135-7
3	54	100.0	53	1	US-08-284-923-2
4	54	100.0	53	4	US-08-619-032B-2
5	54	100.0	53	6	5332669-2
6	44	81.5	48	6	5332669-1
7	44	81.5	754	2	US-08-525-864A-2
8	43	79.6	17	6	5183805-1
9	43	79.6	17	6	5183805-2
10	43	79.6	25	6	5256643-8
11	43	79.6	37	2	US-08-039-364-14
12	43	79.6	37	4	US-08-158-710-14
13	43	79.6	42	1	US-08-168-091A-44
14	43	79.6	44	1	US-08-278-089A-24
15	43	79.6	46	3	US-08-899-437-12
16	43	79.6	46	4	US-09-126-121-12
17	43	79.6	47	3	US-08-753-007A-17
18	43	79.6	47	4	US-09-398-496-17
19	43	79.6	48	4	US-09-020-880-15
20	43	79.6	48	4	US-08-915-096A-13
21	43	79.6	48	4	US-09-101-544-15
22	43	79.6	48	6	5434135-3
23	43	79.6	51	6	517197-50
24	43	79.6	53	1	US-07-869-176-1
25	43	79.6	53	1	US-08-284-923-1
26	43	79.6	53	1	US-08-360-841-2
27	43	79.6	53	2	US-08-861-000-1

28	43	79.6	53	4	US-08-619-032B-1	Sequence 1, Appli
29	43	79.6	53	6	5332669-3	Patent No. 5332669
30	43	79.6	53	6	5434135-2	Patent No. 5434135
31	43	79.6	55	6	5218093-1	Patent No. 5218093
32	43	79.6	88	2	US-07-885-089B-13	Sequence 13, Appli
33	43	79.6	91	1	US-07-847-743B-15	Sequence 15, Appli
34	43	79.6	91	1	US-08-456-201-15	Sequence 15, Appli
35	43	79.6	91	2	US-08-330-161-13	Sequence 13, Appli
36	43	79.6	91	2	US-08-456-241-15	Sequence 15, Appli
37	43	79.6	91	2	US-08-440-401-13	Sequence 13, Appli
38	43	79.6	91	2	US-08-419-878B-13	Sequence 13, Appli
39	43	79.6	91	4	US-09-173-480-13	Sequence 13, Appli
40	43	79.6	91	5	PCT-US92-04295A-15	Sequence 15, Appli
41	43	79.6	93	1	US-08-343-401A-4	Sequence 4, Appli
42	43	79.6	93	1	US-08-445-265A-2	Sequence 2, Appli
43	43	79.6	93	3	US-08-990-442-2	Sequence 2, Appli
44	43	79.6	132	3	US-08-468-846-13	Sequence 13, Appli
45	43	79.6	293	4	US-08-438-745-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-08-597-545-7  
; Sequence 7, Application US/08597545  
; Patent No. 5580738  
; GENERAL INFORMATION:  
; APPLICANT: LABORDA, Jorge  
; TITLE OF INVENTION: Delta-Like Gene Expressed In  
; TITLE OF INVENTION: Neuroendocrine Tumors  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Hardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/597,545  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/989,537  
; FILING DATE: 11-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 40399/166 NIHD  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-597-545-7

Query Match 100.0%; Score 54; DB 1; Length 12;  
Best Local Similarity 90.0%; Pred. No. 0.014;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGXSGDRC 10

Db 3 CVIGXSGDRC 12

## RESULT 2

US-08-457-135-7  
; Sequence 7, Application US/08457135  
; Patent No. 5644031  
; GENERAL INFORMATION:  
; APPLICANT: LABORDA, Jorge  
; TITLE OF INVENTION: Delta-Like Gene Expressed In  
; TITLE OF INVENTION: Neuroendocrine Tumors  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/457,135  
; FILING DATE: 01-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/989,537  
; FILING DATE: 11-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 40399/304/NIHD  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-457-135-7

Query Match 100.0%; Score 54; DB 1; Length 12;  
Best Local Similarity 90.0%; Pred. No. 0.014;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CVIGXSGDRC 10  
Db 3 CVIGYSGDRC 12

## RESULT 3

US-08-284-923-2  
; Sequence 2, Application US/08284923  
; Patent No. 5547935  
; GENERAL INFORMATION:  
; APPLICANT: Mullenbach, Guy T  
; APPLICANT: Blaney, Jeffrey M  
; APPLICANT: Rosenberg, Steven  
; TITLE OF INVENTION: Mutens of Epidermal Growth Factor  
; TITLE OF INVENTION: exhibiting enhanced binding at low ph  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/284,923  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/746,651  
; FILING DATE: 16-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 231.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-601-2708  
; TELEFAX: 510-655-3542  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 53 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-284-923-2

Query Match 100.0%; Score 54; DB 1; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.056;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CVIGXSGDRC 10  
Db 33 CVIGYSGDRC 42

## RESULT 4

US-08-619-032B-2  
; Sequence 2, Application US/08619032B  
; Patent No. 6191106  
; GENERAL INFORMATION:  
; APPLICANT: Mullenbach, Guy T.  
; APPLICANT: Blaney, Jeffrey M.  
; APPLICANT: Rosenberg, Steven  
; TITLE OF INVENTION: MUTENS OF EPIDERMAL GROWTH FACTOR  
; TITLE OF INVENTION: EXHIBITING ENHANCED BINDING AT LOW PH  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: Intellectual Property R-440, P.O. Box 8097  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94662-8097  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/619,032B  
; FILING DATE: 20-MAR-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Guth, Joseph H.  
; REGISTRATION NUMBER: 31,261  
; REFERENCE/DOCKET NUMBER: 0231.004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 923-3888  
; TELEFAX: (510) 655-3542  
; TELEX: N/A  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 53 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single

; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-619-032B-2

Query Match 100.0%; Score 54; DB 4; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.056;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
| | | | | | | | | |  
DB 33 CVIGYSGDRC 42

## RESULT 5

5332669-2  
; Patent No. 5332669  
; APPLICANT: DEUEL, THOMAS F.  
; TITLE OF INVENTION: PROSTATE-DERIVED MITOGEN  
; NUMBER OF SEQUENCES: 3  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/356,739  
; FILING DATE: 24-MAR-1989  
; SEQ ID NO:2;  
; LENGTH: 53  
5332669-2

Query Match 100.0%; Score 54; DB 6; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.056;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
| | | | | | | | | |  
DB 33 CVIGYSGDRC 42

## RESULT 6

5332669-1  
; Patent No. 5332669  
; APPLICANT: DEUEL, THOMAS F.  
; TITLE OF INVENTION: PROSTATE-DERIVED MITOGEN  
; NUMBER OF SEQUENCES: 3  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/356,739  
; FILING DATE: 24-MAR-1989  
; SEQ ID NO:1;  
; LENGTH: 48  
5332669-1

Query Match 81.5%; Score 44; DB 6; Length 48;  
Best Local Similarity 70.0%; Pred. No. 1.8;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
| | | | | | | | | |  
DB 33 CVIGYIGERC 42

## RESULT 7

US-08-525-864A-2  
; Sequence 2, Application US/08525864A  
; Patent No. 5912326  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Han  
; TITLE OF INVENTION: Cerebellum-derived Growth Factors, and Uses  
; TITLE OF INVENTION: Related thereto  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII (text)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/525,864A  
; FILING DATE: 8-SEP-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kara, Catherine J.  
; REGISTRATION NUMBER: 41,106  
; REFERENCE/DOCKET NUMBER: HUI-017  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 754 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-525-864A-2

Query Match 81.5%; Score 44; DB 2; Length 754;  
Best Local Similarity 60.0%; Pred. No. 24;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
| | | | | | | | | |  
DB 280 CPVGYTGDCR 289

## RESULT 8

5183805-1  
; Patent No. 5183805  
; APPLICANT: LEE, JIN S.; BLICK, MARK  
; TITLE OF INVENTION: BIOACTIVE EGF PEPTIDES FOR  
; PROMOTION OF TISSUE REGENERATION AND CANCER THERAPY  
; NUMBER OF SEQUENCES: 2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/567,407  
; FILING DATE: 13-AUG-1990  
; SEQ ID NO:1;  
; LENGTH: 17  
5183805-1

Query Match 79.6%; Score 43; DB 6; Length 17;  
Best Local Similarity 60.0%; Pred. No. 0.94;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
| | | | | | | | | |  
DB 2 CVVGYIGERC 11

## RESULT 9

5183805-2  
; Patent No. 5183805  
; APPLICANT: LEE, JIN S.; BLICK, MARK  
; TITLE OF INVENTION: BIOACTIVE EGF PEPTIDES FOR  
; PROMOTION OF TISSUE REGENERATION AND CANCER THERAPY  
; NUMBER OF SEQUENCES: 2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/567,407  
; FILING DATE: 13-AUG-1990  
; SEQ ID NO:2;  
; LENGTH: 17  
5183805-2

Query Match 79.6%; Score 43; DB 6; Length 17;  
Best Local Similarity 60.0%; Pred. No. 0.94;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```
QY      1  CVIGXSGDRC 10
      ||:|: |:||
Db      2  CVVGYIGERC 11

RESULT 10
5256643-8
; Patent No. 5256643
; APPLICANT: Persico, Maria G.; Salomon, David S.
; TITLE OF INVENTION: HUMAN CRIPTO PROTEIN
; NUMBER OF SEQUENCES: 18
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/530,165
; FILING DATE: 29-MAY-1990
; SEQ ID NO: 8
; LENGTH: 25
5256643-8

Query Match      79.6%; Score 43; DB 6; Length 25;
Best Local Similarity 60.0%; Pred. No. 1.4;
Matches 6; Conservative 3; Mismatches 1; Indels 1; Gaps 0;

QY      1  CVIGXSGDRC 10
      ||:|: |:||
Db      13 CVVGYIGERC 22

RESULT 11
US-08-039-364-14
; Sequence 14, Application US/08039364
; Patent No. 5811393
; GENERAL INFORMATION:
; APPLICANT: Klagsbrun, Michael
; APPLICANT: Abraham, Judith A.
; APPLICANT: Higashiyama, Shigeki
; APPLICANT: Besner, Gail F.
; TITLE OF INVENTION: HEPARIN BINDING MITOGEN WITH
; TITLE OF INVENTION: HEPARIN BINDING MITOGEN WITH
; TITLE OF INVENTION: HOMOLOGY TO EPIDERMAL GROWTH FACTOR
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/039,364
; FILING DATE: 15 JUN 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/598,082
; FILING DATE: 16 OCT 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 05162/002002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
US-08-039-364-14
```

```
Query Match      79.6%; Score 43; DB 2; Length 37;
Best Local Similarity 60.0%; Pred. No. 2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1  CVIGXSGDRC 10
      ||:|: |:||
Db      28 CVVGYIGERC 37
```

```
RESULT 12
US-09-158-710-14
; Sequence 14, Application US/09158710
; Patent No. 6235884
; GENERAL INFORMATION:
; APPLICANT: Klagsbrun, Michael
; APPLICANT: Abraham, Judith A.
; APPLICANT: Higashiyama, Shigeki
; APPLICANT: Besner, Gail F.
; TITLE OF INVENTION: HEPARIN BINDING MITOGEN WITH HOMOLOGY TO EPIDERMAL
; FILE REFERENCE: 05162/002003
; CURRENT APPLICATION NUMBER: US/09/158,710
; CURRENT FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: US 08/039,364
; EARLIER FILING DATE: 1993-06-15
; EARLIER APPLICATION NUMBER: US 07/598,082
; EARLIER FILING DATE: 1990-10-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-158-710-14
```

```
Query Match      79.6%; Score 43; DB 4; Length 37;
Best Local Similarity 60.0%; Pred. No. 2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1  CVIGXSGDRC 10
      ||:|: |:||
Db      28 CVVGYIGERC 37
```

```
RESULT 13
US-08-168-091A-44
; Sequence 44, Application US/08168091A
; Patent No. 5665862
; GENERAL INFORMATION:
; APPLICANT: Fischbach, Gerald
; APPLICANT: Falls, Douglas R.
; APPLICANT: Rosen, Kenneth M.
; APPLICANT: Corfas, Gabriel
; TITLE OF INVENTION: Neurotrophic Factor
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE AND COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/168,091A
; FILING DATE: 15-DEC-1993
```



CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/953,742  
FILING DATE: 29-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Deconti, Giulio A  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: HMI-002CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-168-091A-44

Query Match 79.6%; Score 43; DB 1; Length 42;  
Best Local Similarity 60.0%; Pred. No. 2.2;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
|||:|:|  
Db 31 CVVGIGERC 40

RESULT 14  
US-08-278-089A-24  
Sequence 24, Application US/08278089A  
Patent No. 5681714  
GENERAL INFORMATION:  
APPLICANT: Breitman, Martin L.  
APPLICANT: Rossant, Janet  
APPLICANT: Dumont, Daniel J.  
APPLICANT: Yamaguchi, Terry P.  
TITLE OF INVENTION: No. 5681714el Receptor Tyrosine Kinase  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hereskin & Parr  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/278,089A  
FILING DATE: 20-JUL-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurdydyk, Linda M.  
REGISTRATION NUMBER: 34,971  
REFERENCE/DOCKET NUMBER: 3153-111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
IMMEDIATE SOURCE:  
CLONE: EGF

US-08-278-089A-24

Query Match 79.6%; Score 43; DB 1; Length 44;  
Best Local Similarity 60.0%; Pred. No. 2.3;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
|||:|:|  
Db 33 CVVGIGERC 42

RESULT 15  
US-08-899-437-12  
Sequence 12, Application US/08899437  
Patent No. 6121415  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao  
TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related  
Ligands and Uses Therefor  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/899,437  
FILING DATE: 24-Jul-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Conley, Deirdre L.  
REGISTRATION NUMBER: 36,487  
REFERENCE/DOCKET NUMBER: P1084R1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-2066  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
FEATURE:  
NAME/KEY: hEGF.egf  
LOCATION: 1-46  
IDENTIFICATION METHOD:  
OTHER INFORMATION:

US-08-899-437-12

Query Match 79.6%; Score 43; DB 3; Length 46;  
Best Local Similarity 60.0%; Pred. No. 2.4;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
|||:|:|  
Db 32 CVVGIGERC 41

Search completed: July 2, 2003, 07:25:43  
Job time : 12 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:28:49 ; Search time 103 Seconds  
(without alignments)  
11.164 Million cell updates/sec

Title: US-09-673-785d-8  
Perfect score: 54  
Sequence: 1 CVIGXSGDRC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA.\*

```
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/FCI_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/FCI_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	54	100.0	53	9	US-10-138-158-19
2	54	100.0	53	9	US-10-150-648B-33
3	54	100.0	145	9	US-10-150-648B-30
4	44	81.5	112	9	US-10-211-994-20
5	44	81.5	298	10	US-09-864-675-4
6	44	81.5	878	9	US-10-157-031-171
7	43	79.6	46	9	US-10-136-573A-12
8	43	79.6	46	9	US-09-877-665-12
9	43	79.6	46	9	US-10-215-862-12
10	43	79.6	46	10	US-09-817-647-12
11	43	79.6	47	12	US-10-096-241-17
12	43	79.6	48	9	US-10-201-945-13
13	43	79.6	53	9	US-09-903-327A-9
14	43	79.6	53	9	US-10-150-648B-35
15	43	79.6	53	9	US-10-211-994-4
16	43	79.6	53	10	US-09-848-664-31
17	43	79.6	58	10	US-09-934-706-3
18	43	79.6	91	9	US-10-022-609-13
19	43	79.6	111	9	US-10-211-994-27

```
20 43 79.6 111 9 US-10-211-994-29 Sequence 29, Appl
21 43 79.6 112 9 US-10-211-994-5 Sequence 5, Appl
22 43 79.6 112 9 US-10-211-994-7 Sequence 7, Appl
23 43 79.6 112 9 US-10-211-994-25 Sequence 25, Appl
24 43 79.6 140 10 US-09-280-030-64 Sequence 64, Appl
25 43 79.6 146 9 US-10-150-648B-32 Sequence 32, Appl
26 43 79.6 400 10 US-09-934-706-5 Sequence 5, Appl
27 43 79.6 493 9 US-09-903-327A-13 Sequence 13, Appl
28 43 79.6 741 10 US-09-925-301-930 Sequence 930, Appl
29 41 75.9 572 9 US-09-900-449A-7 Sequence 7, Appl
30 41 75.9 601 9 US-09-900-449A-5 Sequence 5, Appl
31 41 75.9 639 9 US-09-900-449A-4 Sequence 4, Appl
32 40 74.1 53 9 US-10-150-648B-34 Sequence 34, Appl
33 40 74.1 145 9 US-10-150-648B-31 Sequence 31, Appl
34 40 74.1 176 9 US-10-024-599-4 Sequence 4, Appl
35 40 74.1 1050 9 US-09-796-753-114 Sequence 114, Appl
36 39 72.2 94 10 US-09-764-853-716 Sequence 716, Appl
37 39 72.2 769 9 US-09-984-130-67 Sequence 67, Appl
38 39 72.2 769 9 US-10-097-340-157 Sequence 157, Appl
39 38 70.4 636 9 US-09-796-753-100 Sequence 100, Appl
40 38 70.4 636 9 US-09-796-753-124 Sequence 124, Appl
41 38 70.4 723 10 US-09-893-737-100 Sequence 100, Appl
42 38 70.4 3635 9 US-10-037-182-4 Sequence 4, Appl
43 38 70.4 3635 10 US-09-845-583-2 Sequence 2, Appl
44 38 70.4 3712 9 US-10-108-605-103 Sequence 103, Appl
45 38 70.4 4545 10 US-09-873-403-2 Sequence 2, Appl
```

## ALIGNMENTS

## RESULT 1

```
US-10-138-158-19
; Sequence 19, Application US/10138158
; Publication No. US20030036509A1
; GENERAL INFORMATION:
; APPLICANT: STEM CELL PHARMACEUTICALS, INC.
; APPLICANT: TWARDZIK, Daniel R.
; APPLICANT: PERNET, Andre
; APPLICANT: FELKER, Thomas S.
; APPLICANT: PASKELL, Stefan
; APPLICANT: RENO, John M.
; TITLE OF INVENTION: TGF-alpha POLYPEPTIDES, FUNCTIONAL FRAGMENTS AND METHODS OF
; FILE REFERENCE: STEM110-6
; CURRENT APPLICATION NUMBER: US/10/138.158
; CURRENT FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: US 09/641,587
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/559,248
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 09/459,813
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 09/492,935
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 09/378,567
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-158-19
```

Query Match 100.0%; Score 54; DB 9; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.063; Indels 0;  
Matches 9; Conservative 1; Mismatches 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
|||||  
Db 33 CVIGXSGDRC 42

```

RESULT 2
US-10-150-648B-33
; Sequence 33, Application US/10150648B
; Publication No. US20030059802A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau-Goesseels, Sylvie
; APPLICANT: John, Sushil J.
; APPLICANT: Selinger, Leonard B.
; APPLICANT: Benkel, Bernhard F.
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
; TITLE OF INVENTION: factor
; FILE REFERENCE: 60-01
; CURRENT APPLICATION NUMBER: US/10/150,648B
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/292,136
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a
; TITLE: 128,000-molecular weight protein precursor
; JOURNAL: Nature
; VOLUME: 303
; PAGES: 722-725
; DATE: 1983
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380
; DATABASE ENTRY DATE: 1993-04-27
; RELEVANT RESIDUES: Relevant residues FROM 977 TO 1029
US-10-150-648B-33

Query Match      100.0%; Score 54; DB 9; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.063;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
   ||||:||||
Db 33 CVIGYSGDRC 42

RESULT 3
US-10-150-648B-30
; Sequence 30, Application US/10150648B
; Publication No. US20030059802A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau-Goesseels, Sylvie
; APPLICANT: John, Sushil J.
; APPLICANT: Selinger, Leonard B.
; APPLICANT: Benkel, Bernhard F.
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
; TITLE OF INVENTION: factor
; FILE REFERENCE: 60-01
; CURRENT APPLICATION NUMBER: US/10/150,648B
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/292,136
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a
; TITLE: 128,000-molecular weight protein precursor
; JOURNAL: Nature
; VOLUME: 303
; PAGES: 722-725

```

```

; DATE: 1983
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380
; DATABASE ENTRY DATE: 1993-04-27
; RELEVANT RESIDUES: Relevant residues FROM 919 TO 1063
US-10-150-648B-30

Query Match      100.0%; Score 54; DB 9; Length 145;
Best Local Similarity 90.0%; Pred. No. 0.16;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
   ||||:||||
Db 91 CVIGYSGDRC 100

RESULT 4
US-10-211-994-20
; Sequence 20, Application US/10211994
; Publication No. US20030082201A1
; GENERAL INFORMATION:
; APPLICANT: Rao, M.R.S.
; APPLICANT: Sengupta, Paromita
; APPLICANT: Prasad, Sudhanand
; APPLICANT: Burman, Anand C.
; APPLICANT: Mukherjee, Rama
; APPLICANT: Thomas, Becky
; TITLE OF INVENTION: MULTIVALENT SYNTHETIC VACCINE FOR CANCER
; FILE REFERENCE: U014152-1
; CURRENT APPLICATION NUMBER: US/10/211,994
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/309,975
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide and peptide sequence of multivalent vaccine
US-10-211-994-20

Query Match      81.5%; Score 44; DB 9; Length 112;
Best Local Similarity 70.0%; Pred. No. 5.1;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
   ||||:|:|
Db 92 CVIGYIGERC 101

RESULT 5
US-09-864-675-4
; Sequence 4, Application US/09864675
; Patent No. US20020081286A1
; GENERAL INFORMATION:
; APPLICANT: Marchionni, Mark
; TITLE OF INVENTION: NRG-2 NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: POLYPEPTIDES, AND DIAGNOSTIC AND THERAPEUTIC METHODS
; FILE REFERENCE: 04585/049002
; CURRENT APPLICATION NUMBER: US/09/864,675
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/206,495
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-675-4

Query Match      81.5%; Score 44; DB 10; Length 298;

```

```
Best Local Similarity 60.0%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 280 CPVGYTGDCR 289

RESULT 6
US-10-157-031-171
; Sequence 171, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 171
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-157-031-171

Query Match 81.5%; Score 44; DB 9; Length 878;
Best Local Similarity 70.0%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 526 CLPFGSGDRC 535

RESULT 7
US-10-136-573A-12
; Sequence 12, Application US/10136573A
; Patent No. US20020161200A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie Rose
; APPLICANT: Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and
; FILE REFERENCE: P1084RIC2
; CURRENT APPLICATION NUMBER: US/10/136,573A
; CURRENT FILING DATE: 2002-04-29
; PRIOR APPLICATION NUMBER: US 09/480,977
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: US 08/899,437
; PRIOR FILING DATE: 1997-07-24
; PRIOR APPLICATION NUMBER: US 60/052,019
; PRIOR FILING DATE: 1997-07-09
; NUMBER OF SEQ ID NOS: 23
; SEQ ID NO 12
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-136-573A-12

Query Match 79.6%; Score 43; DB 9; Length 46;
Best Local Similarity 60.0%; Pred. No. 3.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 32 CVVGYIGERC 41

RESULT 8
US-09-877-665-12
; Sequence 12, Application US/09877665
; Patent No. US20020164680A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/877,665
; FILING DATE: 08-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/109,206
; FILING DATE: 30-Jun-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: hEGF.egf
; LOCATION: 1-46
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-877-665-12

Query Match 79.6%; Score 43; DB 9; Length 46;
Best Local Similarity 60.0%; Pred. No. 3.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 32 CVVGYIGERC 41

RESULT 9
US-10-215-862-12
; Sequence 12, Application US/10215862
; Publication No. US20030036166A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie Rose
; APPLICANT: Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and
; FILE REFERENCE: P1084R1D2C1
; CURRENT APPLICATION NUMBER: US/10/215,862
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 09/126,663
; PRIOR FILING DATE: 1998-07-30
```

;; PRIOR APPLICATION NUMBER: US 08/899,437  
;; PRIOR FILING DATE: 1997-07-24  
;; PRIOR APPLICATION NUMBER: US 60/052,019  
;; PRIOR FILING DATE: 1997-07-09  
;; NUMBER OF SEQ ID NOS: 23  
;; SEQ ID NO 12  
;; LENGTH: 46  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-215-862-12

Query Match 79.6%; Score 43; DB 9; Length 46;  
Best Local Similarity 60.0%; Pred. No. 3.2;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
Db 32 CVVGYIGRC 41  
||:|: |||

RESULT 10  
US-09-817-647-12  
;; Sequence 12, Application US/09817647  
;; Patent No. US20020082229A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao  
;; TITLE OF INVENTION: Ligands and Uses Therefor  
;; NUMBER OF SEQUENCES: 23  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genentech, Inc.  
;; STREET: 1 DNA Way  
;; CITY: South San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/817,647  
FILING DATE: 26-Mar-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/107,979  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Conley, Delirdre L.  
REGISTRATION NUMBER: 36,487  
REFERENCE/DOCKET NUMBER: P1084R1-2

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-2066  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear

FEATURE:  
NAME/KEY: hEGF.egf  
LOCATION: 1-46  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-09-817-647-12

Query Match 79.6%; Score 43; DB 10; Length 46;  
Best Local Similarity 60.0%; Pred. No. 3.2;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10

Db 32 CVVGYIGRC 41  
||:|: |||

RESULT 11  
US-10-096-241-17  
;; Sequence 17, Application US/10096241  
;; Patent No. US20020127594A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Gearing, David P.  
;; Busfield, Samantha J.

;; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES  
;; NUMBER OF SEQUENCES: 33  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Fish & Richardson P.C.  
;; STREET: 225 Franklin Street  
;; CITY: Boston  
;; STATE: MA

;; COUNTRY: US  
;; ZIP: 02110-2804  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq Version 2.0

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/10/096,241  
;; FILING DATE: 12-Mar-2002  
;; CLASSIFICATION: <Unknown>

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/699,591  
;; FILING DATE: 19-AUG-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fasse, J. Peter

;; REGISTRATION NUMBER: 32,983  
;; REFERENCE/DOCKET NUMBER: 07334/022001  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617-542-5070  
;; TELEFAX: 617-542-8906

;; TELEX: <Unknown>  
;; INFORMATION FOR SEQ ID NO: 17:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 47 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: not relevant

;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 17:  
US-10-096-241-17

Query Match 79.6%; Score 43; DB 12; Length 47;  
Best Local Similarity 60.0%; Pred. No. 3.3;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
Db 33 CVVGYIGRC 42  
||:|: |||

RESULT 12  
US-10-201-945-13  
;; Sequence 13, Application US/10201945  
;; Publication No. US20020188110A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Meissner, Paul S.  
;; Fuldner, Rebecca A.  
;; Adams, Mark D.

;; TITLE OF INVENTION: Transforming Growth Factor Alpha HI  
;; NUMBER OF SEQUENCES: 15  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Human Genome Sciences, Inc.  
;; STREET: 9410 Key West Avenue

CITY: Rockville  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/201,945  
FILING DATE: 25-Jul-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/471,377  
FILING DATE: 23-Dec-1999  
APPLICATION NUMBER: 08/915,096  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 08/208,008  
FILING DATE: 08-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF110DI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8504  
TELEX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-10-201-945-13  
Query Match 79.6%; Score 43; DB 9; Length 48;  
Best Local Similarity 60.0%; Pred. No. 3.3;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CVIGXSGDRC 10  
DB 28 CVVGYIGERC 37  
RESULT 13  
US-09-903-327A-9  
; Sequence 9, Application US/09903327A  
; Patent No. US20020164333A1  
; GENERAL INFORMATION:  
; APPLICANT: Nemerow, Glen R.  
; TITLE OF INVENTION: BIFUNCTIONAL MOLECULES AND VECTORS COMPLEXED THEREWITH FOR TARGET  
; TITLE OF INVENTION: GENE  
; FILE REFERENCE: 22908-1228  
; CURRENT APPLICATION NUMBER: US/09/903,327A  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 09/613,017  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 53  
; TYPE: PRT  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Epidermal Growth Factor (EGF, mature peptide)  
US-09-903-327A-9

Query Match 79.6%; Score 43; DB 9; Length 53;  
Best Local Similarity 60.0%; Pred. No. 3.7;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CVIGXSGDRC 10  
DB 33 CVVGYIGERC 42  
RESULT 14  
US-10-150-648B-35  
; Sequence 35, Application US/10150648B  
; Publication No. US20030059802A1  
; GENERAL INFORMATION:  
; APPLICANT: Bilodeau-Goesseels, Sylvie  
; APPLICANT: John, Sushil J  
; APPLICANT: Selinger, Leonard B..  
; APPLICANT: Benkel, Bernhard F..  
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth  
; TITLE OF INVENTION: factor  
; FILE REFERENCE: 60-01  
; CURRENT APPLICATION NUMBER: US/10/150,648B  
; CURRENT FILING DATE: 2002-05-17  
; PRIOR APPLICATION NUMBER: 60/292,136  
; PRIOR FILING DATE: 2001-05-18  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 35  
; LENGTH: 53  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; PUBLICATION INFORMATION:  
; AUTHORS: Bell, G.I., Fong, N.M., Stempieu, M.M., Wormsted, M.A.,  
; AUTHORS: Caput, D., Ku, L.L., Urdea, M.S., Rall, L.B. and Sanchez-  
; AUTHORS: Pescador, R..  
; TITLE: Human epidermal growth factor precursor: cDNA sequence,  
; TITLE: expression in vitro and gene organization.  
; JOURNAL: Nucleic Acids Research  
; VOLUME: 14  
; ISSUE: 21  
; PAGES: 8427-8446  
; DATE: 1986  
; DATABASE ENTRY DATE: 1993-04-21  
; RELEVANT RESIDUES: Relevant residues FROM 970 TO 1022  
US-10-150-648B-35  
Query Match 79.6%; Score 43; DB 9; Length 53;  
Best Local Similarity 60.0%; Pred. No. 3.7;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CVIGXSGDRC 10  
DB 33 CVVGYIGERC 42  
RESULT 15  
US-10-211-994-4  
; Sequence 4, Application US/10211994  
; Publication No. US20030082201A1  
; GENERAL INFORMATION:  
; APPLICANT: Rao, M.R.S.  
; APPLICANT: Sengupta, Paromita  
; APPLICANT: Prasad, Sudhanand  
; APPLICANT: Burman, Anand C.  
; APPLICANT: Mukherjee, Rama  
; APPLICANT: Thomas, Becky  
; TITLE OF INVENTION: MULTIVALENT SYNTHETIC VACCINE FOR CANCER  
; FILE REFERENCE: U014152-1  
; CURRENT APPLICATION NUMBER: US/10/211,994  
; CURRENT FILING DATE: 2002-08-02  
; PRIOR APPLICATION NUMBER: 60/309,975  
; PRIOR FILING DATE: 2001-08-03

```
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Epidermal growth factor
US-10-211-994-4

Query Match      79.6%; Score 43; DB 9; Length 53;
Best Local Similarity 60.0%; Pred. No. 3.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CVIGXSGDRC 10
   |||: |||
Db 33 CVVGYIGERC 42

Search completed: July 2, 2003, 07:57:18
Job time : 103 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 12 Seconds  
(without alignments)  
80.112 Million cell updates/sec

Title: US-09-673-785D-8

Perfect score: 54

Sequence: 1 CVIGXSGDRC 10

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	1217	1 EGM5MG	epidermal growth f
2	49	90.7	2139	2 A35672	crumbs protein - f
3	47	87.0	1643	2 T14274	versican precursor
4	47	87.0	3381	2 T42389	versican precursor
5	44	81.5	57	2 P04415	ErbB kinase activa
6	44	81.5	372	2 T29359	hypothetical prote
7	44	81.5	1133	1 EGR7	epidermal growth f
8	43	79.6	102	2 B55885	chondroitin sulfat
9	43	79.6	264	2 T16271	hypothetical prote
10	43	79.6	862	2 S43922	versican - pig-tai
11	43	79.6	1207	1 EGHU	epidermal growth f
12	43	79.6	2409	1 A60979	versican precursor
13	42	77.8	2352	2 T30201	Notch homolog prot
14	42	77.8	2531	2 T31070	notch homolog - se
15	42	77.8	4006	2 T09070	probable tenascin
16	41	75.9	447	2 A39321	mucin - rat (fragm
17	41	75.9	601	2 T48539	probable potassium
18	41	75.9	1064	2 A40136	fibropellin Ia - s
19	40	74.1	53	2 S17294	epidermal growth f
20	40	74.1	57	2 B69300	hypothetical prote
21	40	74.1	601	2 T34396	hypothetical prote
22	40	74.1	1203	2 A49175	Notch B protein -
23	40	74.1	2471	2 A49128	cell-fate determin
24	39	72.2	230	2 A44074	probable EGF-like
25	39	72.2	768	2 B41029	integrin beta-8 ch
26	39	72.2	769	2 A41029	integrin beta-8 ch
27	39	72.2	1111	2 T26972	hypothetical prote
28	39	72.2	1531	2 T42218	slit-1 protein hom
29	39	72.2	1964	2 T09059	notch4 - mouse

30 39 72.2 2397 1 A55535 versican precursor  
31 39 72.2 4307 2 T20721 hypothetical prote  
32 38 70.4 117 2 T38295 hypothetical prote  
33 38 70.4 716 2 T51426 hypothetical prote  
34 38 70.4 1086 2 T05407 hypothetical prote  
35 38 70.4 3106 1 S53868 laminin alpha-2 ch  
36 38 70.4 3635 2 T10053 laminin alpha 5 ch  
37 38 70.4 3672 2 T23433 hypothetical prote  
38 38 70.4 3704 2 T37316 probable laminin a  
39 38 70.4 3712 2 S18253 laminin alpha-1 ch  
40 38 70.4 4544 1 S02392 alpha-2-macroglobu  
41 38 70.4 4545 1 S25111 alpha-2-macroglobu  
42 37 68.5 108 2 A46222 hydrophobin Ccg-2  
43 37 68.5 161 2 S03938 neuroendocrine pro  
44 37 68.5 174 2 B90094 60S ribosomal prot  
45 37 68.5 422 2 S74581 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

##### EGMSG

epidermal growth factor precursor - mouse

N:Alternate names: urogastrone precursor

C:Species: Mus musculus (house mouse)

C>Date: 30-Nov-1980 #sequence.revision 11-Aug-1983 #text\_change 19-Jan-2001

C:Accession: A94272; A93304; A92118; A01387

R:Scott, J.; Urdea, M.; Quiroga, M.; Sanchez-Pescador, R.; Fong, N.; Selby, M.; Ruttenberg, A.

Science 221, 236-240, 1983

A:Title: Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor

A:Reference number: A94272; MUID:83223630; PMID:6602382

A:Accession: A94272

A:Molecule type: mRNA

A:Residues: 1-1217 <SC>

A:Cross-references: GB:J00380; NID:gl92993; PIDN:AAA37539.1; PID:g309210

R:Gray, A.; Dull, T.; Ullrich, A.

Nature 303, 722-725, 1983

A:Title: Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molecular weight protein

A:Reference number: A93304; MUID:83219309; PMID:6304537

A:Accession: A93304

A:Molecule type: mRNA

A:Residues: 1-789, 'Y', '791-1047, 'S', 1049-1168 <GRA>

A:Cross-references: GB:J00380

A:Note: the sequence shown by these authors differs from residues 1134-1168 due to an error in the original publication

R:Savage Jr., C.R.; Inagami, T.; Cohen, S.

J. Biol. Chem. 247, 7612-7621, 1972

A:Title: The primary structure of epidermal growth factor.

A:Reference number: A92118; MUID:73048516; PMID:4636327

A:Accession: A92118

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

F:135-176/Domain: LDL receptor YWTD-containing repeat homology <YW03>  
F:177-217/Domain: LDL receptor YWTD-containing repeat homology <YW04>  
F:218-262/Domain: LDL receptor YWTD-containing repeat homology <YW05>  
F:263-307/Domain: LDL receptor YWTD-containing repeat homology <YW06>  
F:324-360/Domain: EGF homology #status atypical <EG1>  
F:366-401/Domain: EGF homology <EG2>  
F:407-442/Domain: EGF homology <EG3>  
F:445-482/Domain: EGF homology <EG4>  
F:486-529/Region: EGF precursor long repeat  
F:489-529/Domain: LDL receptor YWTD-containing repeat homology <YW07>  
F:530-572/Domain: LDL receptor YWTD-containing repeat homology <YW08>  
F:573-615/Domain: LDL receptor YWTD-containing repeat homology <YW09>  
F:616-659/Domain: LDL receptor YWTD-containing repeat homology <YW10>  
F:660-700/Domain: LDL receptor YWTD-containing repeat homology <YW11>  
F:701-743/Domain: LDL receptor YWTD-containing repeat homology <YW12>  
F:751-786/Domain: EGF homology <EG5>  
F:842-875/Domain: EGF homology <EG6>  
F:881-917/Domain: EGF homology <EG7>  
F:923-958/Domain: EGF homology <EG8>  
F:977-1029/Product: epidermal growth factor #status experimental <EGF>  
F:982-1018/Domain: EGF homology <EG9>  
F:1039-1063/Domain: transmembrane #status predicted <TM>  
F:1064-1217/Domain: intracellular #status predicted <INT>  
F:347-360,366-377,373-386,388-401,407-418,414-427,429-442,445-457,453-467,469-482,751-786  
tus predicted  
F:982-996,990-1007,1009-1018/Disulfide bonds: #status experimental

Query Match 100.0%; Score 54; DB 1; Length 1217;  
Best Local Similarity 90.0%; Pred. No. 0.32;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
|||:|||||  
Db 1009 CVIGXSGDRC 1018

## RESULT 2

A35672  
crumbs protein - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 21-Sep-1990 #sequence\_revision 18-Nov-1992 #text\_change 11-Jan-2000  
C:Accession: A35672  
R:Repass, U.; Theres, C.; Knust, E.  
Cell 61, 787-799, 1990  
A:Title: crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila  
A:Reference number: A35672; MUID:90263104; PMID:2344615  
A:Accession: A35672  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2139 <TEP>  
A:Cross-references: GB:M33753  
A>Note: the authors translated the codon GGC for residue 1928 as Cys, and TAT for residue  
C:Genetics:  
A:Gene: FlyBase:crb  
A:Cross-references: FlyBase:FBgn0000368  
A:Superfamily: unassigned EGF-related proteins; EGF homology  
C:Keywords: transmembrane protein  
F:352-385/Domain: EGF homology <EGX1>  
F:392-424/Domain: EGF homology <EGF1>  
F:691-722/Domain: EGF homology <EGF>  
F:767-799/Domain: EGF homology <EGF3>  
F:1878-1914/Domain: EGF homology <EGX2>

Query Match 90.7%; Score 49; DB 2; Length 2139;  
Best Local Similarity 70.0%; Pred. No. 3.8;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
|:|:|||||  
Db 571 CAVGSGDRC 580

## RESULT 3

T14274  
versican precursor, splice form V2 - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 05-May-2000  
C:Accession: T14274  
R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.  
J. Biol. Chem. 273, 15758-15764, 1998  
A:Title: Versican V2 is a major extracellular matrix component of the mature bovine  
A:Reference number: Z17954; MUID:98288320; PMID:9624174  
A:Accession: T14274  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1643 <SCH>  
A:Cross-references: EMBL:AF060458; NID:g3253303; PID:g3253304; PIDN:AAC24360.1  
A:Experimental source: brain  
C:Keywords: glycoprotein  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-1643/Product: versican, splice form V2 #status predicted <MAT>  
F:57,331,352,817,965,1017,1333,1616,1626/Binding site: carbohydrate (Asn) (covalent)

Query Match 87.0%; Score 47; DB 2; Length 1643;  
Best Local Similarity 80.0%; Pred. No. 6.6;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
||:|||||  
Db 1362 CVPGYSGDRC 1371

## RESULT 4

T42389  
versican precursor, splice form V0 - bovine  
N:Alternate names: chondroitin sulfate proteoglycan  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 05-May-2000  
C:Accession: T42389  
R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.  
J. Biol. Chem. 273, 15758-15764, 1998  
A:Title: Versican V2 is a major extracellular matrix component of the mature bovine  
A:Reference number: Z17954; MUID:98288320; PMID:9624174  
A:Accession: T42389  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-3381 <SCH>  
A:Cross-references: EMBL:AF060456; NID:g32533299; PID:g3253300; PIDN:AAC24358.1  
A:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type 1  
C:Keywords: chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-3381/Product: versican, splice form V0 #status predicted <MAT>  
F:57,331,352,817,965,1017,1333,1393,1437,1463,1653,1974,2045,2074,2103,2263,2290,2336

Query Match 87.0%; Score 47; DB 2; Length 3381;  
Best Local Similarity 80.0%; Pred. No. 13;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
||:|||||  
Db 3100 CVPGYSGDRC 3109

## RESULT 5

PC4415  
Erbb kinase activator beta, brain and thymus - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Dec-1997 #sequence\_revision 10-Dec-1997 #text\_change 02-Aug-2002  
C:Accession: PC4415  
R:Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; Mi  
J. Biochem. 122, 675-680, 1997  
A:Title: A novel brain-derived member of the epidermal growth factor family that in  
A:Reference number: JC5700; MUID:98006324; PMID:9348101  
A:Accession: PC4415  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA

A;Residues: 1-57 <HIG>  
A;Cross-references: DBJ:D89997; NID:g2605633; PIDN:BAA23346.1; PID:g2605634  
A;Experimental source: PC-12 cell  
C;Comment: This protein is a member of the epidermal growth factor family. It is functioning the differentiation of MDA-MB-453 cells.  
C;Superfamily: human ERBB kinase activator alpha, brain and thymus; EGF homology F;1-25/Domain: EGF homology (fragment) <EGF>

Query Match 81.5%; Score 44; DB 2; Length 57;  
Best Local Similarity 60.0%; Pred. No. 1;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
I :|:|:|:|  
Db 16 CPVGYTGDRC 25

RESULT 6  
T29359  
hypothetical protein R05G6.9 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T29359  
R;Murray, J.; Le, T.T.  
Submitted to the EMBL Data Library, May 1996  
A;Description: The sequence of C. elegans cosmid R05G6.  
A;Reference number: 220612  
A;Accession: T29359  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-372 <MUR>  
A;Cross-references: EMBL:U58746; PIDN:AAB00626.1; GSPDB:GN00022; CESP:R05G6.9  
A;Experimental source: strain Bristol N2; clone R05G6  
C;Genetics:  
A;Gene: CESP:R05G6.9  
A;Map position: 4  
A;Introns: 80/1; 161/1; 245/1; 286/1

Query Match 81.5%; Score 44; DB 2; Length 372;  
Best Local Similarity 60.0%; Pred. No. 5.6;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
I :|:|:|:|  
Db 233 CYLGYSGDKC 242

RESULT 7  
EGRT  
epidermal growth factor precursor - rat  
N;Alternate names: urogastrophil precursor  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 31-Dec-1988 #sequence\_revision 14-Aug-1998 #text\_change 18-Jun-1999  
C;Accession: I52995; S05074; S01974; A25425; S18419; S08288  
R;Saggi, S.J.; Safirstein, R.; Price, P.M.  
DNA Cell Biol. 11, 481-487, 1992  
A;Title: Cloning and Sequencing of the Rat Preproepidermal Growth Factor cDNA: Comparison  
A;Reference number: I52995; MUID:92398779; PMID:1524680  
A;Accession: I52995  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1133 <RES>  
A;Cross-references: EMBL:U04842; NID:g440236; PIDN:AAB60436.1; PID:g440237  
R;Simpson, R.  
Submitted to the EMBL Data Library, August 1988  
A;Reference number: S05074  
A;Accession: S05074  
A;Molecule type: mRNA  
A;Residues: 'W', 966, 'RWL', 970-1023, 'NW', 1026-1108, 'SGAGVSSGPQPFVFLV', 1126, 'HQ' <SIM>  
A;Cross-references: EMBL:X12748  
R;Dorow, D.S.; Simpson, R.J.  
Nucleic Acids Res. 16, 9338, 1988  
A;Title: Cloning and sequence analysis of a cDNA for rat epidermal growth factor.

## RESULT 8

B5885  
chondroitin sulfate proteoglycan MV3 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 16-Jul-1999  
C:Accession: B5885  
R:Zako, M.; Shinomura, T.; Ujita, M.; Ito, K.; Kimata, K.  
J. Biol. Chem. 270, 3914-3918, 1995  
A:Title: Expression of PG-M(V3), an alternatively spliced form of PG-M without a chondroitin sulfate domain, in human skin fibroblasts  
A:Reference number: A55885; MUID:95181355; PMID:7876137  
A:Accession: B5885  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-102 <ZAK>  
A:Cross-references: GB:S75879; GB:D32039  
C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EGF homology <EG1>  
F:24-55/Domain: EGF homology <EG1>  
F:62-93/Domain: EGF homology <EG2>

Query Match 79.6%; Score 43; DB 2; Length 102;

Best Local Similarity 70.0%; Pred. No. 2.6;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10

DB 46 CVPGLSGDQC 55

## RESULT 9

T16271  
hypothetical protein F35D2.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T16271  
R:Connell, M.  
submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of C. elegans cosmid F35D2.  
A:Reference number: 218488  
A:Accession: T16271  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-264 <CON>  
A:Cross-references: EMBL:U28741; NID:g861290; PID:g861291; PIDN:AAA68325.1; CESP:F35D2.3  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:F35D2.3  
A:Introns: 40/3; 71/3; 160/3; 197/3

Query Match 79.6%; Score 43; DB 2; Length 264;

Best Local Similarity 70.0%; Pred. No. 6.2;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10

DB 76 CVPGLSGDRC 85

## RESULT 10

S43922  
versican - pig-tailed macaque (fragments)  
N:Alternate names: chondroitin sulfate proteoglycan  
C:Species: Macaca nemestrina (pig-tailed macaque)  
C:Date: 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 16-Jul-1999  
C:Accession: S43922  
R:Iso, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.  
Matrix Biol. 14, 213-225, 1994  
A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by immunohistochemistry  
A:Reference number: S43921; MUID:95005762; PMID:7921538  
A:Accession: S43922  
A:Molecule type: mRNA  
A:Residues: 1-233; 234-525; 526-862 <YAO>  
A:Cross-references: EMBL:S72413

A:Note: 507-Ser was also found

A:Note: the authors translated the codon GCC for residue 50 as Val, AAG for residue 469 as Asn

C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; chondroitin sulfate proteoglycan; extracellular matrix

F:1-37/Domain: link protein repeat homology (fragment) <LNK1>

F:58-139/Domain: link protein repeat homology <LNK1>

F:722-753/Domain: EGF homology <EG1>

F:760-791/Domain: EGF homology <EG2>

Query Match 79.6%; Score 43; DB 2; Length 862;

Best Local Similarity 70.0%; Pred. No. 18;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10

DB 744 CVPGLSGDQC 753

## RESULT 11

EGH9  
epidermal growth factor precursor [validated] - human  
N:Alternate names: urogastrone precursor  
C:Species: Homo sapiens (man)  
C:Date: 30-Nov-1980 #sequence\_revision 14-Aug-1998 #text\_change 08-Dec-2000  
C:Accession: A25531; A01388; A33517; A29721; S45282; S45283  
R:Bell, G.I.; Fong, N.M.; Stempien, M.M.; Wormsted, M.A.; Caput, D.; Ku, L.; Urdea, M.  
Nucleic Acids Res. 14, 8427-8446, 1986  
A:Title: Human epidermal growth factor precursor: cDNA sequence, expression in vitro  
A:Reference number: A25531; MUID:87066721; PMID:3491360  
A:Accession: A25531  
A:Molecule type: mRNA  
A:Residues: 1-1207 <BEL>  
A:Cross-references: EMBL:X04571; NID:g31120; PIDN:CAA28240.1; PID:g31121  
A:Note: 708-Met was also found  
A:Note: intron positions were also determined  
R:Gregory, H.; Preston, B.M.  
Int. J. Pept. Protein Res. 9, 107-118, 1977  
A:Title: The primary structure of human urogastrone.  
A:Reference number: A01388; MUID:77117897; PMID:300079  
A:Accession: A01388  
A:Molecule type: protein  
A:Residues: 971-1023 <GRE>  
A:Note: some of the molecules lack Arg-1023  
R:Furuya, M.; Akashi, S.; Hirayama, K.  
Biochem. Biophys. Res. Commun. 163, 1100-1106, 1989  
A:Title: The primary structure of human EGF produced by genetic engineering, studied  
A:Reference number: A33517; MUID:89391964; PMID:2789514  
A:Accession: A33517  
A:Molecule type: protein  
A:Residues: 971-1023 <FUR>  
R:Tsukumo, K.; Nakamura, H.; Sakamoto, S.  
Biochem. Biophys. Res. Commun. 145, 126-133, 1987  
A:Title: Purification and characterization of high molecular weight human epidermal  
A:Reference number: A29721; MUID:87241488; PMID:3297054  
A:Accession: A29721  
A:Molecule type: protein  
A:Residues: 829-834, 'X', 836-839, 'X', 841-845, 'X', 847-848 <TSU>  
A:Note: this is the amino-terminal sequence of a high molecular weight form of EGF.  
R:Svoboda, M.; Bauhofer, A.; Schwind, P.; Bade, E.; Rasched, I.; Przybylski, M.  
Biochim. Biophys. Acta 1206, 35-41, 1994  
A:Title: Structural characterization and biological activity of recombinant human epi  
A:Reference number: S45282; MUID:94242778; PMID:8186248  
A:Accession: S45282  
A:Molecule type: protein  
A:Residues: 'M', 971-1023 <SVO>  
A:Note: expressed recombinant protein  
A:Accession: S45283  
A:Molecule type: protein  
A:Residues: 'MKYP', 970-1023 <SV2>  
A:Note: expressed recombinant protein  
C:Comment: Epidermal growth factor (EGF) stimulates the proliferation and differentiation of gastrointestinal cell proliferation.

C:Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in some C:Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound protein  
C:Genetics:  
A:Gene: GDB:EGF  
A:Cross-references: GDB:119105; OMIM:131530  
A:Map position: 4q25-q425  
A:Introns: 43/1; 109/3; 170/2; 246/2; 314/1; 356/3; 397/1; 438/1; 480/1; 525/3; 575/2; 6  
A:Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-contain  
C:Keywords: duplication; growth factor; tandem repeat; transmembrane protein  
F:1-22/Domain: signal sequence status predicted <SIG>  
F:23-1207/Product: epidermal growth factor proprotein, membrane-bound form #status predi  
F:23-1032/Domain: extracellular #status predicted <EXT>  
F:43-479/Region: EGF precursor long repeat <LR1>  
F:46-85/Domain: LDL receptor YWTD-containing repeat homology <YW01>  
F:86-127/Domain: LDL receptor YWTD-containing repeat homology <YW02>  
F:128-169/Domain: LDL receptor YWTD-containing repeat homology <YW03>  
F:170-211/Domain: LDL receptor YWTD-containing repeat homology <YW04>  
F:212-256/Domain: LDL receptor YWTD-containing repeat homology <YW05>  
F:257-301/Domain: LDL receptor YWTD-containing repeat homology <YW06>  
F:318-354/Domain: EGF homology <EG1>  
F:360-395/Domain: EGF homology <EG2>  
F:401-436/Domain: EGF homology <EG3>  
F:439-476/Domain: EGF homology <EG4>  
F:480-523/Region: EGF precursor long repeat <LR2>  
F:483-523/Domain: LDL receptor YWTD-containing repeat homology <YW07>  
F:524-566/Domain: LDL receptor YWTD-containing repeat homology <YW08>  
F:567-609/Domain: LDL receptor YWTD-containing repeat homology <YW09>  
F:610-653/Domain: LDL receptor YWTD-containing repeat homology <YW10>  
F:654-694/Domain: LDL receptor YWTD-containing repeat homology <YW11>  
F:695-737/Domain: LDL receptor YWTD-containing repeat homology <YW12>  
F:745-780/Domain: EGF homology <EG5>  
F:835-868/Domain: EGF homology <EG6>  
F:874-910/Domain: EGF homology <EG7>  
F:916-951/Domain: EGF homology <EG8>  
F:971-1023/Product: epidermal growth factor #status experimental <EGF>  
F:976-1012/Domain: EGF homology <EG9>  
F:1033-1057/Domain: transmembrane #status predicted <TM>  
F:1058-1207/Domain: intracellular #status predicted <INT>  
F:318-330,325-339,341-354,360-371,367-380,382-395,401-412,408-421,423-436,439-451,447-448  
fide bonds: #status predicted  
F:976-990,984-1001,1003-1012/Disulfide bonds: #status experimental

Query Match 79.6%; Score 43; DB 1; Length 1207;  
Best Local Similarity 60.0%; Pred. No. 24;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
DB 1003 CVVGYIGERC 1012

RESULT 12  
A60979  
versican precursor - human  
N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteoglycan  
N:Contains: glial hyaluronate-binding protein  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 19-Jan-2001  
C:Accession: S06014; S43921; A60979; A30358; A29348; A45131; I54179  
EMBO J. 8, 2975-2981, 1989  
R:Zimmermann, D.R.; Ruoslahti, E.  
A:Title: Multiple domains of the large fibroblast proteoglycan, versican.  
A:Reference number: S06014; MUID:90059882; PMID:2583089  
A:Accession: S06014  
A:Molecule type: mRNA  
A:Residues: 1-2409 <ZIM>  
A:Cross-references: GB:X15998; NID:937662; PIDN:CAA34128.1; PID:g37663  
R:Yao, L.Y.; Moody, C.; Schoenher, E.; Wight, T.N.; Sandell, L.J.  
Matrix Biol. 14, 213-225, 1994  
A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by  
A:Reference number: S43921; MUID:95005762; PMID:7921538  
A:Accession: S43921  
A:Molecule type: mRNA

A:Residues: 208-440;1094-1385;1910-2246 <YAO>  
R:Bignami, A.; Lane, W.S.; Andrews, D.; Dahl, D.  
Brain Res. Bull. 22, 67-70, 1989  
A:Title: Structural similarity of hyaluronate binding proteins in brain and cartilag  
A:Reference number: A60979; MUID:89229983; PMID:2469524  
A:Accession: A60979  
A:Molecule type: protein  
A:Residues: 171-210;289-303 <BIG>  
R:Perides, G.; Lane, W.S.; Andrews, D.; Dahl, D.; Bignami, A.  
J. Biol. Chem. 264, 5981-5987, 1989  
A:Title: Isolation and partial characterization of a glial hyaluronate-binding prote  
A:Reference number: A30358; MUID:89174663; PMID:2466833  
A:Accession: A30358  
A:Molecule type: protein  
A:Residues: 24-50;80-87,'D','89-119;128-155;167-218;229-259,'IR';261-268;277-283,'G'  
R:Krusius, T.; Gehlsen, K.R.; Ruoslahti, E.  
J. Biol. Chem. 267, 13120-13125, 1992  
A:Title: A fibroblast chondroitin sulfate proteoglycan core protein contains lectin-  
A:Reference number: A29348; MUID:88007514; PMID:2820964  
A:Accession: A29348  
A:Molecule type: mRNA  
A:Residues: 1725,'V',1727-2409 <KRU>  
A:Cross-references: GB:J02814  
R:Perides, G.; Rahemtulla, F.; Lane, W.S.; Asher, R.A.; Bignami, A.  
J. Biol. Chem. 267, 23883-23887, 1992  
A:Title: Isolation of a large aggregating proteoglycan from human brain.  
A:Reference number: A45131; MUID:93054750; PMID:1429726  
A:Contents: brain  
A:Accession: A45131  
A:Molecule type: protein  
A:Residues: 21-22,'X',24-37 <PE2>  
A:Experimental source: brain  
A:Note: Sequence extracted from NCBI backbone (NCBIP:118884)  
R:Iozzo, R.V.; Naso, M.F.; Cannizzaro, L.A.; Wasmuth, J.J.; McPherson, J.D.  
Genomics 14, 845-851, 1992  
A:Title: Mapping of the versican proteoglycan gene (CSPG2) to the long arm of human  
A:Reference number: I54179; MUID:93122792; PMID:1478664  
A:Accession: I54179  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 251-347 <RES>  
A:Cross-references: GB:SS2488; NID:g263313; PIDN:AAB24878.1; PID:g263314  
C:Genetics:  
A:Gene: GDB:CSPG2  
A:Cross-references: GDB:I27873; OMIM:118661  
A:Map position: 5q12-5q14  
C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-2409/Product: proteoglycan 24K core protein #status predicted <MAT>  
F:167-244/Domain: link protein repeat homology <LNK1>  
F:265-346/Domain: link protein repeat homology <LNK2>  
F:559-1654/Domain: chondroitin sulfate attachment #status predicted <GAG>  
F:2106-2137/Domain: EGF homology <EG1>  
F:2144-2175/Domain: EGF homology <EG2>  
F:2182-2302/Domain: C-type lectin homology <LCH>  
F:2309-2365/Domain: complement factor H repeat homology <FHD>

Query Match 79.6%; Score 43; DB 1; Length 2409;  
Best Local Similarity 70.0%; Pred. No. 45;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10  
DB 2128 CVPGYSGDQC 2137

RESULT 13  
T30201  
Notch homolog protein - sea squirt (Halocynthia roretzi)  
C:Species: Halocynthia roretzi  
C:Date: 02-Sep-2000 #sequence revision 02-Sep-2000 #text change 02-Sep-2000  
A:Accession: T30201  
R:Hori, S.; Saitoh, T.; Matsumoto, M.; Makabe, K.W.; Nishida, H.

Dev. Genes Evol. 207, 371-380, 1997

A:Title: Notch Homologue from *Halocynthia roretzi* is preferentially expressed in the cerebellum  
A:Reference number: Z20775  
A:Accession: T30201  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2352 <HOR>  
A:Cross-references: EMBL:AB001327; NID:dl204472; PID:dl026501; PIDN:BAA25571.1  
C:Genetics:  
A:Gene: Notch

## Query Match

Best Local Similarity 77.8%; Score 42; DB 2; Length 2352;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10

Db 367 CVAGYSGPRC 376

## RESULT 14

T31070

notch homolog - sea urchin (*Lytechinus variegatus*)  
C:Species: *Lytechinus variegatus* (variegated urchin)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T31070  
R:Sherwood, D.R.; McClay, D.R.  
Development 124, 3363-3374, 1997  
A:Title: Identification and localization of a sea urchin Notch homologue: insights into the evolution of the Notch signaling pathway  
A:Reference number: Z20966; MUID:97454256; PMID:9310331  
A:Accession: T31070  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2531 <SHE>  
A:Cross-references: EMBL:AF000634; NID:g2570350; PID:g2570351; PIDN:AAB82088.1  
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

## Query Match

Best Local Similarity 77.8%; Score 42; DB 2; Length 2531;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10

Db 615 CPVGTSGDNC 624

## RESULT 15

T09070

probable tenascin X - mouse  
C:Species: *Mus musculus* (house mouse)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jan-2000  
C:Accession: T09070  
R:Roven, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, C.; Suck, D.  
submitted to the EMBL Data Library, October 1997  
A:Description: Sequence of the mouse major histocompatibility locus class III region.  
A:Reference number: Z16543  
A:Accession: T09070  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4006 <ROW>  
A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958  
C:Genetics:  
A:Gene: TNX  
A:Map position: 17  
A:Introns: 124/1; 735/1; 773/3; 826/1; 914/1; 1037/1; 1135/1; 1232/1; 1329/1; 1440/1; 1501/1; 3115/1; 3208/1; 3305/1; 3405/1; 3517/1; 3558/1; 3606/1; 3646/1; 3694/1; 3737/3; 3773/3  
C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin type III  
C:Keywords: extracellular matrix  
F:422-448/Domain: EGF homology <EGF>  
F:826-906/Domain: fibronectin type III repeat homology <3FR>  
F:3789-3997/Domain: fibrinogen beta/gamma homology <FBG>

## Query Match

Best Local Similarity 77.8%; Score 42; DB 2; Length 4006;

Best Local Similarity 60.0%;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10

Db 532 CAVGSGDDC 541

Search completed: July 2, 2003, 07:26:26  
Job time: 14 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:02 ; Search time 10.3333 Seconds  
(without alignments)  
40.138 Million cell updates/sec

Title: US-09-673-785D-8  
Perfect score: 54  
Sequence: 1 CVIGXSGDRC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	54	100.0	1217	1	EGF_MOUSE
2	49	90.7	2139	1	CRB_DROME
3	47	87.0	3381	1	PGCV_BOVIN
4	44	81.5	756	1	NRG2_MOUSE
5	44	81.5	1133	1	EGF_RAT
6	43	79.6	862	1	PGCV_MACNE
7	43	79.6	1207	1	EGF_HUMAN
8	43	79.6	3396	1	PGCV_HUMAN
9	42	77.8	1242	1	JAG1_BRARE
10	41	75.9	1064	1	FBP1_STRPU
11	41	75.9	2920	1	CLR2_MOUSE
12	40	74.1	53	1	EGF_PIG
13	40	74.1	57	1	Y402_ARCFU
14	40	74.1	2470	1	NTC2_MOUSE
15	40	74.1	2471	1	NTC2_HUMAN
16	40	74.1	2471	1	NTC2_RAT
17	39	72.2	183	1	YRF3_SHIFL
18	39	72.2	230	1	SPT_DROME
19	39	72.2	768	1	ITB8_RABIT
20	39	72.2	769	1	ITB8_HUMAN
21	39	72.2	1964	1	NTC4_MOUSE
22	39	72.2	2738	1	PGCV_RAT
23	39	72.2	3358	1	PGCV_MOUSE
24	38	70.4	117	1	YEP2_SCHPO
25	38	70.4	3106	1	LMA2_MOUSE
26	38	70.4	3672	1	LML2_CAEEL
27	38	70.4	3712	1	LMA_DROME
28	38	70.4	3718	1	LMA5_MOUSE
29	38	70.4	4544	1	LRPL_HUMAN
30	37	68.5	108	1	RODL_NEUCR
31	37	68.5	161	1	7B2_XENLA
32	37	68.5	490	1	MURE_YERPE
33	37	68.5	494	1	MURE_ECO57

RESULT 1  
EGF\_MOUSE  
ID EGF\_MOUSE STANDARD; PRT; 1217 AA.  
AC P01132; 1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 21-JUL-1986 (Rel. 01, Last annotation update)  
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal growth factor].  
DE EGF.  
GN EGF.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=63223630; PubMed=6602382;  
RA Scott J., Urdea M., Quiroga M., Sanchez-Pescador R., Pong N.M., Selby M., Rutter W.J., Bell G.I.;  
RT "Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins.";  
RL Science 221:236-240(1983).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83219309; PubMed=6304537;  
RA Gray A., Dull T.J., Ullrich A.;  
RT "Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molecular weight protein precursor.";  
RL Nature 303:722-725(1983).  
RN [3]  
RP SEQUENCE OF 977-1029.  
RX MEDLINE=73048516; PubMed=4636327;  
RA Savage C.R. Jr., Inagami T., Cohen S.;  
RT "The primary structure of epidermal growth factor.";  
RL J. Biol. Chem. 247:7612-7621(1972).  
RN [4]  
RP DISULFIDE BONDS.  
RX MEDLINE=74025498; PubMed=4750422;  
RA Savage C.R. Jr., Hash J.H., Cohen S.;  
RT "Epidermal growth factor. Location of disulfide bonds.";  
RL J. Biol. Chem. 248:7669-7672(1973).  
RN [5]  
RP STRUCTURE BY NMR OF 977-1029.  
RX MEDLINE=92118798; PubMed=1731873;  
RA Montellione G.T., Wüthrich K., Burgess A.W., Nice E.C., Wagner G., Gibson K.D., Scheraga H.A.;  
RT "Solution structure of murine epidermal growth factor determined by NMR spectroscopy and refined by energy minimization with restraints.";  
RL Biochemistry 31:236-249(1992).  
RN [6]  
RP STRUCTURE BY NMR OF 977-1029.  
RX MEDLINE=93075811; PubMed=1445923;  
RA Konda D., Inagami F.;  
RT "Three-dimensional nuclear magnetic resonance structures of mouse epidermal growth factor in acidic and physiological pH solutions.";  
RN

## ALIGNMENTS

34 37 68.5 494 1 MURE\_ECOLI P22188 escherichia  
35 37 68.5 494 1 MURE\_PASMU P57815 pasteurella  
36 37 68.5 495 1 MURE\_SALTI Q829h3 salmonella  
37 37 68.5 495 1 MURE\_SALTY Q8zu7 salmonella  
38 37 68.5 496 1 MURE\_ANASP Q8ywf0 anabaena sp  
39 37 68.5 670 1 DYLL\_HUMAN P54792 homo sapien  
40 37 68.5 675 1 PRKS\_MOUSE Q03761 mus musculus  
41 37 68.5 799 1 TRKA\_RAT P35739 rattus norv  
42 37 68.5 1213 1 JAG3\_BRARE Q90y54 brachydanio  
43 37 68.5 2003 1 NTC4\_HUMAN Q94466 homo sapien  
44 37 68.5 2319 1 NTC3\_RAT Q91172 rattus norv  
45 37 68.5 2437 1 NTC1\_BRARE P46530 brachydanio

Biochemistry 31:11928-11939(1992).  
[7]  
RN STRUCTURE BY NMR OF 980-1024.  
RX MEDLINE=99180407; PubMed=10082370;  
RA Barnham K.J., Torres A.M., Alewood D., Alewood P.F., Domagala T.,  
RA Nice E.C., Norton R.S.;  
RT "Role of the 6-20 disulfide bridge in the structure and activity of  
epidermal growth factor.";  
RL Protein Sci. 7:1738-1749(1998).  
CC -1- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS  
CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME  
CC FIBROBLASTS IN CELL CULTURE.  
CC -1- SUBCELLULAR LOCATION: type I membrane protein.  
CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.  
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 1134  
CC TO 1168 DUE TO A FRAMESHIFT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J00380; AAA37539.1; -  
DR EMBL: V00741; CAA24115.1; ALT\_FRAME.  
DR EMBL: V00741; CAA24115.1; -  
DR FIR: A01387; EGMSMG.  
DR PDB: 1EGF; 31-JAN-94.  
DR PDB: 3EGF; 31-JAN-94.  
DR PDB: 1EPG; 31-JAN-94.  
DR PDB: 1EPH; 31-JAN-94.  
DR PDB: 1EPI; 31-JAN-94.  
DR PDB: 1EPJ; 31-JAN-94.  
DR PDB: 1EPJ; 31-JAN-94.  
DR MGD: MGI:95290; Egf.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001336; EGF\_1.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR000033; Ldl\_receptor\_rep.  
DR Pfam: PF00008; EGF; 8.  
DR Pfam: PF00058; ldl\_recept\_b; 7.  
DR PRINTS: PR00009; EGF\_TGF.  
DR SMART: SM00179; EGF\_Ca; 2.  
DR SMART: SM00001; EGF\_like; 7.  
DR SMART: SM00135; LY; 9.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 3.  
DR PROSITE: PS00022; EGF\_1; 1.  
DR PROSITE: PS01186; EGF\_2; 6.  
DR PROSITE: PS01187; EGF\_Ca; 3.  
KW EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;  
KW Signal; 3D-structure  
FT SIGNAL 1  
FT CHAIN 29 1217  
FT CHAIN 977 1029  
FT DOMAIN 29 1038  
FT TRANSMEM 1039 1058  
FT DOMAIN 1059 1217  
FT DOMAIN 327 361  
FT DOMAIN 362 402  
FT DOMAIN 403 443  
FT DOMAIN 441 483  
FT DOMAIN 747 787  
FT DOMAIN 838 876  
FT DOMAIN 877 918  
FT DOMAIN 919 959  
FT DOMAIN 978 1019  
FT DISULFID 366 377  
FT DISULFID 373 386  
FT DISULFID 388 401  
FT DISULFID 407 418  
POTENTIAL.  
PRO-EPIDERMAL GROWTH FACTOR.  
EPIDERMAL GROWTH FACTOR.  
EXTRACELLULAR (POTENTIAL).  
POTENTIAL.  
CYTOPLASMIC (POTENTIAL).  
EGF-LIKE 1 (INCOMPLETE).  
EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
EGF-LIKE 3.  
EGF-LIKE 4.  
EGF-LIKE 5.  
EGF-LIKE 6.  
EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).  
EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).  
EGF-LIKE 9.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.

FT DISULFID 414 427  
FT DISULFID 429 442  
FT DISULFID 445 457  
FT DISULFID 453 467  
FT DISULFID 469 482  
FT DISULFID 751 762  
FT DISULFID 758 771  
FT DISULFID 773 786  
FT DISULFID 842 853  
FT DISULFID 847 862  
FT DISULFID 864 875  
FT DISULFID 881 895  
FT DISULFID 888 904  
FT DISULFID 906 917  
FT DISULFID 923 936  
FT DISULFID 930 945  
FT DISULFID 947 958  
FT DISULFID 982 996  
FT DISULFID 990 1007  
FT DISULFID 1009 1018  
FT DOMAIN 1024 1029  
FT CARBOHYD 111 111  
FT CARBOHYD 410 410  
FT CARBOHYD 810 810  
FT CARBOHYD 944 944  
FT CONFLICT 790 790  
FT CONFLICT 1048 1048  
FT STRAND 995 997  
FT STRAND 1006 1008  
FT STRAND 1010 1010  
FT TURN 1011 1012  
FT STRAND 1013 1014  
FT STRAND 1020 1021  
SQ SEQUENCE 1217 AA; 133144 MW; A9C7F3D512F82873 CRC64;  
Query Match 100.0%; Score 54; DB 1; Length 1217;  
Best Local Similarity 90.0%; Pred. No. 0.05;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CVIGXSGDRC 10  
Db 1009 CVIGXSGDRC 1018  
||||:||||  
CRB\_DROME STANDARD; PRT; 2139 AA.  
ID CRB\_DROME  
AC P10040;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Crumbs protein precursor (95F).  
GN CRB.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Oregon-R; TISSUE=Embryo;  
RX MEDLINE=90263104; PubMed=2344615;  
RA Tepass U., Theres C., Knust E.;  
RT "Crumbs encodes an EGF-like protein expressed on apical membranes of  
RT Drosophila epithelial cells and required for organization of  
RT epithelia.";  
RL Cell 61:787-799(1990).  
RN [2]  
RP SEQUENCE OF 1663-1955 FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=87218537; PubMed=3107986;  
RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,

NOT REQUIRED FOR FULL BIOLOGICAL ACTIVITY  
IN VIVO.  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
D -> Y (IN REF. 2).  
A -> S (IN REF. 2).



RA Vaessin H., Campos-Ortega J.A.;  
 RT "EGF homologous sequences encoded in the genome of Drosophila  
 RL melanogaster, and their relation to neurogenic genes."; EMBO J. 6:761-766(1987).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,  
 CC POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL  
 CC POLARITY. IT MAY ACT AS A SIGNAL.  
 CC -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).  
 CC -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M33753; AAA28428.1; ALT\_SEQ.  
 DR EMBL; X05144; CAA28793.1; --  
 DR PIR; B26637; B26637.  
 DR PIR; A35672; A35672.  
 DR HSP; P00740; IEDM.  
 DR FlyBase; FBgn000368; crb.  
 DR InterPro; IPR00152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001861; EGF\_Ca.  
 DR InterPro; IPR001438; EGF\_II.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR Pfam; PF00008; EGF; 26.  
 DR Pfam; PF00054; laminin\_G; 3.  
 DR PRINTS; PR00010; EGFBLD.  
 DR SMART; SM00179; EGF\_CA; 11.  
 DR SMART; SM00001; EGF\_like; 16.  
 DR SMART; SM00282; LamG; 3.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 15.  
 DR PROSITE; PS00022; EGF\_1; 26.  
 DR PROSITE; PS01186; EGF\_2; 17.  
 DR PROSITE; PS01187; EGF\_CA; 12.  
 DR PROSITE; PS00025; LAM\_G\_DOMAIN; 3.  
 KW Differentiation; Repeat; EGF-like domain; Transmembrane;  
 KW Glycoprotein; Signal; Phosphorylation.  
 FT SIGNAL 1 90  
 FT CHAIN 91 2139 CRUMBS PROTEIN.  
 FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 2085 2111 POTENTIAL.  
 FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 267 303 EGF-LIKE 1.  
 FT DOMAIN 306 343 EGF-LIKE 2.  
 FT DOMAIN 348 386 EGF-LIKE 3.  
 FT DOMAIN 388 425 EGF-LIKE 4.  
 FT DOMAIN 427 463 EGF-LIKE 5.  
 FT DOMAIN 464 500 EGF-LIKE 6.  
 FT DOMAIN 501 532 EGF-LIKE 7.  
 FT DOMAIN 545 581 EGF-LIKE 8.  
 FT DOMAIN 582 611 EGF-LIKE 9.  
 FT DOMAIN 609 646 EGF-LIKE 10.  
 FT DOMAIN 648 685 EGF-LIKE 11.  
 FT DOMAIN 687 723 EGF-LIKE 12.  
 FT DOMAIN 725 761 EGF-LIKE 13.  
 FT DOMAIN 763 800 EGF-LIKE 14.  
 FT DOMAIN 802 838 EGF-LIKE 15.  
 FT DOMAIN 840 902 EGF-LIKE 16.  
 FT DOMAIN 904 940 EGF-LIKE 17.  
 FT DOMAIN 942 978 EGF-LIKE 18.  
 FT DOMAIN 980 1021 EGF-LIKE 19.  
 FT DOMAIN 1023 1205 LAMININ G-LIKE 1.  
 FT DOMAIN 1207 1243 EGF-LIKE 20.  
 FT DOMAIN 1250 1480 LAMININ G-LIKE 2.  
 FT DOMAIN 1481 1517 EGF-LIKE 21.  
 FT DOMAIN 1518 1557 CRUMBS PROTEIN.  
 FT DOMAIN 1558 1594 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1595 1621 POTENTIAL.  
 FT DOMAIN 1622 1658 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 1659 1695 EGF-LIKE 1.  
 FT DOMAIN 1696 1732 EGF-LIKE 2.  
 FT DOMAIN 1733 1769 EGF-LIKE 3.  
 FT DOMAIN 1770 1806 EGF-LIKE 4.  
 FT DOMAIN 1807 1843 EGF-LIKE 5.  
 FT DOMAIN 1844 1880 EGF-LIKE 6.  
 FT DOMAIN 1881 1917 EGF-LIKE 7.  
 FT DOMAIN 1918 1954 EGF-LIKE 8.  
 FT DOMAIN 1955 1991 EGF-LIKE 9.  
 FT DOMAIN 1992 2028 EGF-LIKE 10.  
 FT DOMAIN 2029 2065 EGF-LIKE 11.  
 FT DOMAIN 2066 2102 EGF-LIKE 12.  
 FT DOMAIN 2103 2139 EGF-LIKE 13.  
 FT DOMAIN 2140 2176 EGF-LIKE 14.  
 FT DOMAIN 2177 2213 EGF-LIKE 15.  
 FT DOMAIN 2214 2250 EGF-LIKE 16.  
 FT DOMAIN 2251 2287 EGF-LIKE 17.  
 FT DOMAIN 2288 2324 EGF-LIKE 18.  
 FT DOMAIN 2325 2361 EGF-LIKE 19.  
 FT DOMAIN 2362 2398 EGF-LIKE 20.  
 FT DOMAIN 2399 2435 EGF-LIKE 21.  
 FT DOMAIN 2436 2472 EGF-LIKE 22.  
 FT DOMAIN 2473 2509 EGF-LIKE 23.  
 FT DOMAIN 2510 2546 EGF-LIKE 24.  
 FT DOMAIN 2547 2583 EGF-LIKE 25.  
 FT DOMAIN 2584 2620 EGF-LIKE 26.  
 FT DOMAIN 2621 2657 EGF-LIKE 27.  
 FT DOMAIN 2658 2694 EGF-LIKE 28.  
 FT DOMAIN 2695 2731 EGF-LIKE 29.  
 FT DOMAIN 2732 2768 EGF-LIKE 30.  
 FT DOMAIN 2769 2805 EGF-LIKE 31.  
 FT DOMAIN 2806 2842 EGF-LIKE 32.  
 FT DOMAIN 2843 2879 EGF-LIKE 33.  
 FT DOMAIN 2880 2916 EGF-LIKE 34.  
 FT DOMAIN 2917 2953 EGF-LIKE 35.  
 FT DOMAIN 2954 2990 EGF-LIKE 36.  
 FT DOMAIN 2991 3027 EGF-LIKE 37.  
 FT DOMAIN 3028 3064 EGF-LIKE 38.  
 FT DOMAIN 3065 3101 EGF-LIKE 39.  
 FT DOMAIN 3102 3138 EGF-LIKE 40.  
 FT DOMAIN 3139 3175 EGF-LIKE 41.  
 FT DOMAIN 3176 3212 EGF-LIKE 42.  
 FT DOMAIN 3213 3249 EGF-LIKE 43.  
 FT DOMAIN 3250 3286 EGF-LIKE 44.  
 FT DOMAIN 3287 3323 EGF-LIKE 45.  
 FT DOMAIN 3324 3360 EGF-LIKE 46.  
 FT DOMAIN 3361 3397 EGF-LIKE 47.  
 FT DOMAIN 3398 3434 EGF-LIKE 48.  
 FT DOMAIN 3435 3471 EGF-LIKE 49.  
 FT DOMAIN 3472 3508 EGF-LIKE 50.  
 FT DOMAIN 3509 3545 EGF-LIKE 51.  
 FT DOMAIN 3546 3582 EGF-LIKE 52.  
 FT DOMAIN 3583 3619 EGF-LIKE 53.  
 FT DOMAIN 3620 3656 EGF-LIKE 54.  
 FT DOMAIN 3657 3693 EGF-LIKE 55.  
 FT DOMAIN 3694 3730 EGF-LIKE 56.  
 FT DOMAIN 3731 3767 EGF-LIKE 57.  
 FT DOMAIN 3768 3804 EGF-LIKE 58.  
 FT DOMAIN 3805 3841 EGF-LIKE 59.  
 FT DOMAIN 3842 3878 EGF-LIKE 60.  
 FT DOMAIN 3879 3915 EGF-LIKE 61.  
 FT DOMAIN 3916 3952 EGF-LIKE 62.  
 FT DOMAIN 3953 3989 EGF-LIKE 63.  
 FT DOMAIN 3990 4026 EGF-LIKE 64.  
 FT DOMAIN 4027 4063 EGF-LIKE 65.  
 FT DOMAIN 4064 4100 EGF-LIKE 66.  
 FT DOMAIN 4101 4137 EGF-LIKE 67.  
 FT DOMAIN 4138 4174 EGF-LIKE 68.  
 FT DOMAIN 4175 4211 EGF-LIKE 69.  
 FT DOMAIN 4212 4248 EGF-LIKE 70.  
 FT DOMAIN 4249 4285 EGF-LIKE 71.  
 FT DOMAIN 4286 4322 EGF-LIKE 72.  
 FT DOMAIN 4323 4359 EGF-LIKE 73.  
 FT DOMAIN 4360 4396 EGF-LIKE 74.  
 FT DOMAIN 4397 4433 EGF-LIKE 75.  
 FT DOMAIN 4434 4470 EGF-LIKE 76.  
 FT DOMAIN 4471 4507 EGF-LIKE 77.  
 FT DOMAIN 4508 4544 EGF-LIKE 78.  
 FT DOMAIN 4545 4581 EGF-LIKE 79.  
 FT DOMAIN 4582 4618 EGF-LIKE 80.  
 FT DOMAIN 4619 4655 EGF-LIKE 81.  
 FT DOMAIN 4656 4692 EGF-LIKE 82.  
 FT DOMAIN 4693 4729 EGF-LIKE 83.  
 FT DOMAIN 4730 4766 EGF-LIKE 84.  
 FT DOMAIN 4767 4803 EGF-LIKE 85.  
 FT DOMAIN 4804 4840 EGF-LIKE 86.  
 FT DOMAIN 4841 4877 EGF-LIKE 87.  
 FT DOMAIN 4878 4914 EGF-LIKE 88.  
 FT DOMAIN 4915 4951 EGF-LIKE 89.  
 FT DOMAIN 4952 4988 EGF-LIKE 90.  
 FT DOMAIN 4989 5025 EGF-LIKE 91.  
 FT DOMAIN 5026 5062 EGF-LIKE 92.  
 FT DOMAIN 5063 5099 EGF-LIKE 93.  
 FT DOMAIN 5100 5136 EGF-LIKE 94.  
 FT DOMAIN 5137 5173 EGF-LIKE 95.  
 FT DOMAIN 5174 5210 EGF-LIKE 96.  
 FT DOMAIN 5211 5247 EGF-LIKE 97.  
 FT DOMAIN 5248 5284 EGF-LIKE 98.  
 FT DOMAIN 5285 5321 EGF-LIKE 99.  
 FT DOMAIN 5322 5358 EGF-LIKE 100.  
 FT DOMAIN 5359 5395 EGF-LIKE 101.  
 FT DOMAIN 5396 5432 EGF-LIKE 102.  
 FT DOMAIN 5433 5469 EGF-LIKE 103.  
 FT DOMAIN 5470 5506 EGF-LIKE 104.  
 FT DOMAIN 5507 5543 EGF-LIKE 105.  
 FT DOMAIN 5544 5580 EGF-LIKE 106.  
 FT DOMAIN 5581 5617 EGF-LIKE 107.  
 FT DOMAIN 5618 5654 EGF-LIKE 108.  
 FT DOMAIN 5655 5691 EGF-LIKE 109.  
 FT DOMAIN 5692 5728 EGF-LIKE 110.  
 FT DOMAIN 5729 5765 EGF-LIKE 111.  
 FT DOMAIN 5766 5802 EGF-LIKE 112.  
 FT DOMAIN 5803 5839 EGF-LIKE 113.  
 FT DOMAIN 5840 5876 EGF-LIKE 114.  
 FT DOMAIN 5877 5913 EGF-LIKE 115.  
 FT DOMAIN 5914 5950 EGF-LIKE 116.  
 FT DOMAIN 5951 5987 EGF-LIKE 117.  
 FT DOMAIN 5988 6024 EGF-LIKE 118.  
 FT DOMAIN 6025 6061 EGF-LIKE 119.  
 FT DOMAIN 6062 6098 EGF-LIKE 120.  
 FT DOMAIN 6099 6135 EGF-LIKE 121.  
 FT DOMAIN 6136 6172 EGF-LIKE 122.  
 FT DOMAIN 6173 6209 EGF-LIKE 123.  
 FT DOMAIN 6210 6246 EGF-LIKE 124.  
 FT DOMAIN 6247 6283 EGF-LIKE 125.  
 FT DOMAIN 6284 6320 EGF-LIKE 126.  
 FT DOMAIN 6321 6357 EGF-LIKE 127.  
 FT DOMAIN 6358 6394 EGF-LIKE 128.  
 FT DOMAIN 6395 6431 EGF-LIKE 129.  
 FT DOMAIN 6432 6468 EGF-LIKE 130.  
 FT DOMAIN 6469 6505 EGF-LIKE 131.  
 FT DOMAIN 6506 6542 EGF-LIKE 132.  
 FT DOMAIN 6543 6579 EGF-LIKE 133.  
 FT DOMAIN 6580 6616 EGF-LIKE 134.  
 FT DOMAIN 6617 6653 EGF-LIKE 135.  
 FT DOMAIN 6654 6690 EGF-LIKE 136.  
 FT DOMAIN 6691 6727 EGF-LIKE 137.  
 FT DOMAIN 6728 6764 EGF-LIKE 138.  
 FT DOMAIN 6765 6801 EGF-LIKE 139.  
 FT DOMAIN 6802 6838 EGF-LIKE 140.  
 FT DOMAIN 6839 6875 EGF-LIKE 141.  
 FT DOMAIN 6876 6912 EGF-LIKE 142.  
 FT DOMAIN 6913 6949 EGF-LIKE 143.  
 FT DOMAIN 6950 6986 EGF-LIKE 144.  
 FT DOMAIN 6987 7023 EGF-LIKE 145.  
 FT DOMAIN 7024 7060 EGF-LIKE 146.  
 FT DOMAIN 7061 7097 EGF-LIKE 147.  
 FT DOMAIN 7098 7134 EGF-LIKE 148.  
 FT DOMAIN 7135 7171 EGF-LIKE 149.  
 FT DOMAIN 7172 7208 EGF-LIKE 150.  
 FT DOMAIN 7209 7245 EGF-LIKE 151.  
 FT DOMAIN 7246 7282 EGF-LIKE 152.  
 FT DOMAIN 7283 7319 EGF-LIKE 153.  
 FT DOMAIN 7320 7356 EGF-LIKE 154.  
 FT DOMAIN 7357 7393 EGF-LIKE 155.  
 FT DOMAIN 7394 7430 EGF-LIKE 156.  
 FT DOMAIN 7431 7467 EGF-LIKE 157.  
 FT DOMAIN 7468 7504 EGF-LIKE 158.  
 FT DOMAIN 7505 7541 EGF-LIKE 159.  
 FT DOMAIN 7542 7578 EGF-LIKE 160.  
 FT DOMAIN 7579 7615 EGF-LIKE 161.  
 FT DOMAIN 7616 7652 EGF-LIKE 162.  
 FT DOMAIN 7653 7689 EGF-LIKE 163.  
 FT DOMAIN 7690 7726 EGF-LIKE 164.  
 FT DOMAIN 7727 7763 EGF-LIKE 165.  
 FT DOMAIN 7764 7800 EGF-LIKE 166.  
 FT DOMAIN 7801 7837 EGF-LIKE 167.  
 FT DOMAIN 7838 7874 EGF-LIKE 168.  
 FT DOMAIN 7875 7911 EGF-LIKE 169.  
 FT DOMAIN 7912 7948 EGF-LIKE 170.  
 FT DOMAIN 7949 7985 EGF-LIKE 171.  
 FT DOMAIN 7986 8022 EGF-LIKE 172.  
 FT DOMAIN 8023 8059 EGF-LIKE 173.  
 FT DOMAIN 8060 8096 EGF-LIKE 174.  
 FT DOMAIN 8097 8133 EGF-LIKE 175.  
 FT DOMAIN 8134 8170 EGF-LIKE 176.  
 FT DOMAIN 8171 8207 EGF-LIKE 177.  
 FT DOMAIN 8208 8244 EGF-LIKE 178.  
 FT DOMAIN 8245 8281 EGF-LIKE 179.  
 FT DOMAIN 8282 8318 EGF-LIKE 180.  
 FT DOMAIN 8319 8355 EGF-LIKE 181.  
 FT DOMAIN 8356 8392 EGF-LIKE 182.  
 FT DOMAIN 8393 8429 EGF-LIKE 183.  
 FT DOMAIN 8430 8466 EGF-LIKE 184.  
 FT DOMAIN 8467 8503 EGF-LIKE 185.  
 FT DOMAIN 8504 8540 EGF-LIKE 186.  
 FT DOMAIN 8541 8577 EGF-LIKE 187.  
 FT DOMAIN 8578 8614 EGF-LIKE 188.  
 FT DOMAIN 8615 8651 EGF-LIKE 189.  
 FT DOMAIN 8652 8688 EGF-LIKE 190.  
 FT DOMAIN 8689 8725 EGF-LIKE 191.  
 FT DOMAIN 8726 8762 EGF-LIKE 192.  
 FT DOMAIN 8763 8799 EGF-LIKE 193.  
 FT DOMAIN 8800 8836 EGF-LIKE 194.  
 FT DOMAIN 8837 8873 EGF-LIKE 195.  
 FT DOMAIN 8874 8910 EGF-LIKE 196.  
 FT DOMAIN 8911 8947 EGF-LIKE 197.  
 FT DOMAIN 8948 8984 EGF-LIKE 198.  
 FT DOMAIN 8985 9021 EGF-LIKE 199.  
 FT DOMAIN 9022 9058 EGF-LIKE 200.  
 FT DOMAIN 9059 9095 EGF-LIKE 201.  
 FT DOMAIN 9096 9132 EGF-LIKE 202.  
 FT DOMAIN 9133 9169 EGF-LIKE 203.  
 FT DOMAIN 9170 9206 EGF-LIKE 204.  
 FT DOMAIN 9207 9243 EGF-LIKE 205.  
 FT DOMAIN 9244 9280 EGF-LIKE 206.  
 FT DOMAIN 9281 9317 EGF-LIKE 207.  
 FT DOMAIN 9318 9354 EGF-LIKE 208.  
 FT DOMAIN 9355 9391 EGF-LIKE 209.  
 FT DOMAIN 9392 9428 EGF-LIKE 210.  
 FT DOMAIN 9429 9465 EGF-LIKE 211.  
 FT DOMAIN 9466 9502 EGF-LIKE 212.  
 FT DOMAIN 9503 9539 EGF-LIKE 213.  
 FT DOMAIN 9540 9576 EGF-LIKE 214.  
 FT DOMAIN 9577 9613 EGF-LIKE 215.  
 FT DOMAIN 9614 9650 EGF-LIKE 216.  
 FT DOMAIN 9651 9687 EGF-LIKE 217.  
 FT DOMAIN 9688 9724 EGF-LIKE 218.  
 FT DOMAIN 9725 9761 EGF-LIKE 219.  
 FT DOMAIN 9762 9798 EGF-LIKE 220.  
 FT DOMAIN 9799 9835 EGF-LIKE 221.  
 FT DOMAIN 9836 9872 EGF-LIKE 222.  
 FT DOMAIN 9873 9909 EGF-LIKE 223.  
 FT DOMAIN 9910 9946 EGF-LIKE 224.  
 FT DOMAIN 9947 9983 EGF-LIKE 225.  
 FT DOMAIN 9984 10000 EGF-LIKE 226.  
 FT DOMAIN 10001 10017 EGF-LIKE 227.  
 FT DOMAIN 10018 10034 EGF-LIKE 228.  
 FT DOMAIN 10035 10051 EGF-LIKE 229.  
 FT DOMAIN 10052 10068 EGF-LIKE 230.  
 FT DOMAIN 10069 10085 EGF-LIKE 231.  
 FT DOMAIN 10086 10102 EGF-LIKE 232.  
 FT DOMAIN 10103 10119 EGF-LIKE 233.  
 FT DOMAIN 10120 10136 EGF-LIKE 234.  
 FT DOMAIN 10137 10153 EGF-LIKE 235.  
 FT DOMAIN 10154 10170 EGF-LIKE 236.  
 FT DOMAIN 10171 10187 EGF-LIKE 237.  
 FT DOMAIN 10188 10204 EGF-LIKE 238.  
 FT DOMAIN 10205 10221 EGF-LIKE 239.  
 FT DOMAIN 10222 10238 EGF-LIKE 240.  
 FT DOMAIN 10239 10255 EGF-LIKE 241.  
 FT DOMAIN 10256 10272 EGF-LIKE 242.  
 FT DOMAIN 10273 10289 EGF-LIKE 243.  
 FT DOMAIN 10290 10306 EGF-LIKE 244.  
 FT DOMAIN 10307 10323 EGF-LIKE 245.  
 FT DOMAIN 10324 10340 EGF-LIKE 246.  
 FT DOMAIN 10341 10357 EGF-LIKE 247.  
 FT DOMAIN 10358 10374 EGF-LIKE 248.  
 FT DOMAIN 10375 10391 EGF-LIKE 249.  
 FT DOMAIN 10392 10408 EGF-LIKE 250.  
 FT DOMAIN 10409 10425 EGF-LIKE 251.  
 FT DOMAIN 10426 10442 EGF-LIKE 252.  
 FT DOMAIN 10443 10459 EGF-LIKE 253.  
 FT DOMAIN 10460 10476 EGF-LIKE 254.  
 FT DOMAIN 10477 10493 EGF-LIKE 255.  
 FT DOMAIN 10494 10510 EGF-LIKE 256.  
 FT DOMAIN 10511 10527 EGF-LIKE 257.  
 FT DOMAIN 10528 10544 EGF-LIKE 258.  
 FT DOMAIN 10545 10561 EGF-LIKE 259.  
 FT DOMAIN 10562 10578 EGF-LIKE 260.  
 FT DOMAIN 10579 10595 EGF-LIKE 261.  
 FT DOMAIN 10596 10612 EGF-LIKE 262.  
 FT DOMAIN 10613 10629 EGF-LIKE 263.  
 FT DOMAIN 10630 10646 EGF-LIKE 264.  
 FT DOMAIN 10647 10663 EGF-LIKE 265.  
 FT DOMAIN 10664 10680 EGF-LIKE 266.  
 FT DOMAIN 10681 10697 EGF-LIKE 267.  
 FT DOMAIN 10698 10713 EGF-LIKE 268.  
 FT DOMAIN 10714 10730 EGF-LIKE 269.  
 FT DOMAIN 10731 10747 EGF-LIKE 270.  
 FT DOMAIN 10748 10764 EGF-LIKE 271.  
 FT DOMAIN 10765 10781 EGF-LIKE 272.  
 FT DOMAIN 10782 10798 EGF-LIKE 273.  
 FT DOMAIN 10799 10815 EGF-LIKE 274.  
 FT DOMAIN 10816 10832 EGF-LIKE 275.  
 FT DOMAIN 10833 10849 EGF-LIKE 276.  
 FT DOMAIN 10850 10866 EGF-LIKE 277.  
 FT DOMAIN 10867 10883 EGF-LIKE 278.  
 FT DOMAIN 10884 10900 EGF-LIKE 279.  
 FT DOMAIN 10901 10917 EGF-LIKE 280.  
 FT DOMAIN 10918 10934 EGF-LIKE 281.  
 FT DOMAIN 10935 10951 EGF-LIKE 282.  
 FT DOMAIN 10952 10968 EGF-LIKE 283.  
 FT DOMAIN 10969 10985 EGF-LIKE 284.  
 FT DOMAIN 10986 11002 EGF-LIKE 285.  
 FT DOMAIN 11003 11018 EGF-LIKE 286.  
 FT DOMAIN 11019 11035 EGF-LIKE 287.  
 FT DOMAIN 11036 11052 EGF-LIKE 288.  
 FT DOMAIN 11053 11069 EGF-LIKE 289.  
 FT DOMAIN 11070 11086 EGF-LIKE 290.  
 FT DOMAIN 11087 11103 EGF-LIKE 291.  
 FT DOMAIN 11104 11120 EGF-LIKE 292.  
 FT DOMAIN 11121 11137 EGF-LIKE 293.  
 FT DOMAIN 11138 11154 EGF-LIKE 294.  
 FT DOMAIN 11155 11171 EGF-LIKE 295.  
 FT DOMAIN 11172 11188 EGF-LIKE 296.  
 FT DOMAIN 11189 11205 EGF-LIKE 297.  
 FT DOMAIN 11206 11222 EGF-LIKE 298.  
 FT DOMAIN 11223 11239 EGF-LIKE 299.  
 FT DOMAIN 11240 11256 EGF-LIKE 300.  
 FT DOMAIN 11257 11273 EGF-LIKE 301.  
 FT DOMAIN 11274 11290 EGF-LIKE 302.  
 FT DOMAIN 11291 11307 EGF-LIKE 303.  
 FT DOMAIN 11308 11324 EGF-LIKE 304.  
 FT DOMAIN 11325 11341 EGF-LIKE 305.  
 FT DOMAIN 11342 11358 EGF-LIKE 306.  
 FT DOMAIN 11359 11375 EGF-LIKE 307.  
 FT DOMAIN 11376 11391 EGF-LIKE 308.  
 FT DOMAIN 11392 11407 EGF-LIKE 309.  
 FT DOMAIN 11408 11423 EGF-LIKE 310.  
 FT DOMAIN 11424 11440 EGF-LIKE 311.  
 FT DOMAIN 11441 11457 EGF-LIKE 312.  
 FT DOMAIN 11458 11473 EGF-LIKE 313.  
 FT DOMAIN 11474 11490 EGF-LIKE 314.  
 FT DOMAIN 11491 11507 EGF-LIKE 315.  
 FT DOMAIN 11508 11524 EGF-LIKE 316.  
 FT DOMAIN 11525 11541 EGF-LIKE 317.  
 FT DOMAIN 11542 11558 EGF-LIKE 318.  
 FT DOMAIN 11559 11575 EGF-LIKE 319.  
 FT DOMAIN 11576 11592 EGF-LIKE 320.  
 FT DOMAIN 11593 11609 EGF-LIKE 321.  
 FT DOMAIN 11610 11626 EGF-LIKE 322.  
 FT DOMAIN 11627 11643 EGF-LIKE 323.  
 FT DOMAIN 11644 11660 EGF-LIKE 324.  
 FT DOMAIN 11661 11677 EGF-LIKE 325.  
 FT DOMAIN 11678 11694 EGF-LIKE 326.  
 FT DOMAIN 11695 11710 EGF-LIKE 327.  
 FT DOMAIN 11711 11727 EGF-LIKE 328.  
 FT DOMAIN 11728 11744 EGF-LIKE 329.  
 FT DOMAIN 11745 11761 EGF-LIKE 330.  
 FT DOMAIN 11762 11778 EGF-LIKE 331.  
 FT DOMAIN 11779 11795 EGF-LIKE 332.  
 FT DOMAIN 11796 11812 EGF-LIKE 333.  
 FT DOMAIN 11813 11829 EGF-LIKE 334.  
 FT DOMAIN 11830 11846 EGF-LIKE 335.  
 FT DOMAIN 11847 11863 EGF-LIKE 336.  
 FT DOMAIN 11864 11880 EGF-LIKE 337.  
 FT DOMAIN 11881 11897 EGF-LIKE 338.  
 FT DOMAIN 11898 11914 EGF-LIKE 339.  
 FT DOMAIN 11915 11930 EGF-LIKE 340.  
 FT DOMAIN 11931 11947 EGF-LIKE 341.  
 FT DOMAIN 11948 11964 EGF-LIKE 342.  
 FT DOMAIN 11965 11981 EGF-LIKE 343.  
 FT DOMAIN 11982 12000 EGF-LIKE 344.  
 FT DOMAIN 12001 12017 EGF-LIKE 345.  
 FT DOMAIN 12018 12034 EGF-LIKE 346.  
 FT DOMAIN 12035 12051 EGF-LIKE 347.  
 FT DOMAIN 12052 12068 EGF-LIKE 348.  
 FT DOMAIN 12069 12085 EGF-LIKE 349.  
 FT DOMAIN 12086 12102 EGF-LIKE 350.  
 FT DOMAIN 12103 12119 EGF-LIKE 351.  
 FT DOMAIN 12120 12136 EGF-LIKE 352.  
 FT DOMAIN 12137 12153 EGF-LIKE 353.  
 FT DOMAIN 12154 12170 EGF-LIKE 354.  
 FT DOMAIN 12171 12187 EGF-LIKE 355.  
 FT DOMAIN 12188 12204 EGF-LIKE 356.  
 FT DOMAIN 12205 12221 EGF-LIKE 357.  
 FT DOMAIN 12222 12238 EGF-LIKE 358.  
 FT DOMAIN 12239 12255 EGF-LIKE 359.  
 FT DOMAIN 12256 12272 EGF-LIKE 360.  
 FT DOMAIN 12273 12289 EGF-LIKE 361.  
 FT DOMAIN 12290 12306 EGF-LIKE 362.  
 FT DOMAIN 12307 12323 EGF-LIKE 363.  
 FT DOMAIN 12324 12340 EGF-LIKE 364.  
 FT DOMAIN 12341 12357 EGF-LIKE 365.  
 FT DOMAIN 12358 12374 EGF-LIKE 366.  
 FT DOMAIN 12375 12391 EGF-LIKE 367.  
 FT DOMAIN 12392 12408 EGF-LIKE 368.  
 FT DOMAIN 12409 12425 EGF-LIKE 369.  
 FT DOMAIN 12426 12442 EGF-LIKE 370.  
 FT DOMAIN 12443 12459 EGF-LIKE 371.  
 FT DOMAIN 12460 12476 EGF-LIKE 372.  
 FT DOMAIN 12477 12493 EGF-LIKE 373.  
 FT DOMAIN 12494 12510 EGF-LIKE 374.  
 FT DOMAIN 12511 12527 EGF-LIKE 375.  
 FT DOMAIN 12528 12544 EGF-LIKE 376.  
 FT DOMAIN 12545 12561 EGF-LIKE 377.  
 FT DOMAIN 12562 12578 EGF-LIKE 378.  
 FT DOMAIN 12579 12595 EGF-LIKE 379.  
 FT DOMAIN 12596 12612 EGF-LIKE 380.  
 FT DOMAIN 12613 12629 EGF-LIKE 381.  
 FT DOMAIN 12630 12646 EGF-LIKE 382.  
 FT DOMAIN 12647 12663 EGF-LIKE 383.  
 FT DOMAIN 12664 12680 EGF-LIKE 384.  
 FT DOMAIN 12681 12697 EGF-LIKE 385.  
 FT DOMAIN 12698 12714 EGF-LIKE 386.  
 FT DOMAIN 12715 12731 EGF-LIKE 387.  
 FT DOMAIN 12732 12748 EGF-LIKE 388.  
 FT DOMAIN 12749 12765 EGF-LIKE 389.  
 FT DOMAIN 12766 12782 EGF-LIKE 390.  
 FT DOMAIN 12783 12800 EGF-LIKE 391.  
 FT DOMAIN 12801 12817 EGF-LIKE 392.  
 FT DOMAIN 12818 12834 EGF-LIKE 393.  
 FT DOMAIN 12835 12851 EGF-LIKE 394.  
 FT DOMAIN 12852 12868 EGF-LIKE 395.  
 FT DOMAIN 12869 12885 EGF-LIKE 396.  
 FT DOMAIN 12886 12902 EGF-LIKE 397.  
 FT DOMAIN 12903 12919 EGF-LIKE 398.  
 FT DOMAIN 12920 12936 EGF-LIKE 399.  
 FT DOMAIN 12937 12953 EGF-LIKE 400.  
 FT DOMAIN 12954 12970 EGF-LIKE 401.  
 FT DOMAIN 12971 12987 EGF-LIKE 402.  
 FT DOMAIN 12988 13004 EGF-LIKE 403.  
 FT DOMAIN 13005 13020 EGF-LIKE 404.  
 FT DOMAIN 13021 13037 EGF-LIKE 405.  
 FT DOMAIN 13038 13054 EGF-LIKE 406.  
 FT DOMAIN 13055 13070 EGF-LIKE 407.  
 FT DOMAIN 13071 13087 EGF-LIKE 408.  
 FT DOMAIN 13088 13104 EGF-LIKE 409.  
 FT DOMAIN 13105 13121 EGF-LIKE 410.  
 FT DOMAIN 13122 13138 EGF-LIKE 411.  
 FT DOMAIN 13139 13155 E

FT DISULFID 1768 1783 BY SIMILARITY.  
 FT DISULFID 1785 1794 BY SIMILARITY.  
 FT DISULFID 1801 1812 BY SIMILARITY.  
 FT DISULFID 1806 1821 BY SIMILARITY.  
 FT DISULFID 1823 1832 BY SIMILARITY.  
 FT DISULFID 1839 1850 BY SIMILARITY.  
 FT DISULFID 1844 1859 BY SIMILARITY.  
 FT DISULFID 1861 1870 BY SIMILARITY.  
 FT DISULFID 1883 1893 BY SIMILARITY.  
 FT DISULFID 1905 1914 BY SIMILARITY.  
 FT DISULFID 1919 1930 BY SIMILARITY.  
 FT DISULFID 1924 1939 BY SIMILARITY.  
 FT DISULFID 1941 1950 BY SIMILARITY.  
 FT DISULFID 1957 1968 BY SIMILARITY.  
 FT DISULFID 1962 1977 BY SIMILARITY.  
 FT DISULFID 1979 1988 BY SIMILARITY.  
 FT DISULFID 1995 2008 BY SIMILARITY.  
 FT DISULFID 2002 2017 BY SIMILARITY.  
 FT DISULFID 2019 2028 BY SIMILARITY.  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).  
  
 Query Match 90.7%; Score 49; DB 1; Length 2139;  
 Best Local Similarity 70.0%; Pred. No. 0.72;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 CVIGXSGDRC 10  
 Db 571 CAVGSGDRC 580  
  
 RESULT 3  
 ID PGCV\_BOVIN STANDARD; PRT; 3381 AA.  
 AC P81282; 077609; 077610; 077611; 077612;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Versican core protein precursor (large fibroblast proteoglycan)  
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial  
 hyaluronate-binding protein) (GHAP).  
 GN CPBG2.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RN SEQUENCE FROM N.A. (ISOFORMS VO; V1; V2 AND V3).  
 RC TISSUE=Forebrain;  
 RA MEDLINE=9628320; PubMed=9624174;  
 RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,  
 RA Zimmermann D.R.;  
 RT "Versican V2 is a major extracellular matrix component of the mature  
 bovine brain."  
 RL J. Biol. Chem. 273:15758-15764(1998).  
 RN [2]  
 RN SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331  
 AND 342-348.  
 RC TISSUE=Spinal cord;  
 RA MEDLINE=92062692; PubMed=1720020;  
 RA Perides G., Biviano F., Bignami A.;  
 RT "Interaction of a brain extracellular matrix protein with hyaluronid  
 acid."  
 RL Biochim. Biophys. Acta 1075:248-258(1991).  
 CC -1- FUNCTION: May play a role in intercellular signaling and in

connecting cells with the extracellular matrix. May take part in  
 the regulation of cell motility, growth and differentiation. Binds  
 hyaluronic acid.  
 -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
 -1- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1, V2  
 and V3; are produced by alternative splicing.  
 -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed  
 in the central nervous system, and in a number of mesenchymal and  
 epithelial tissues; the major isoform V2 is restricted to the  
 central nervous system.  
 -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development  
 (By similarity).  
 -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
 -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.  
 -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; AF060456; AAC24358.1; -;  
 EMBL; AF060457; AAC24359.1; -;  
 EMBL; AF060458; AAC24360.1; -;  
 EMBL; AF060459; AAC24361.1; -;  
 HSSP; P01132; 1EPG.  
 InterPro; IPR000152; Asx\_hydroxyl.  
 InterPro; IPR000561; EGF-like.  
 InterPro; IPR000742; EGF\_2.  
 InterPro; IPR001881; EGF\_Ca.  
 InterPro; IPR003599; Ig.  
 InterPro; IPR003600; Ig\_MHC.  
 InterPro; IPR003600; Ig\_Like.  
 InterPro; IPR001304; Lectin\_C.  
 InterPro; IPR000538; Link.  
 InterPro; IPR000436; Sush1\_SCR\_CCP.  
 Pfam; PF00008; EGF; 2.  
 Pfam; PF00047; Ig; 1.  
 Pfam; PF00059; lectin\_c; 1.  
 Pfam; PF00084; sush1; 1.  
 Pfam; PF00193; Xlink; 2.  
 PRINTS; PR01265; LINKMODULE.  
 ProDom; PD000918; Link; 1.  
 SMART; SM00032; CCP; 1.  
 SMART; SM00034; CLECT; 1.  
 SMART; SM00181; EGF; 2.  
 SMART; SM00179; EGF\_CA; 2.  
 SMART; SM00409; IG; 1.  
 SMART; SM00410; IG\_Like; 1.  
 SMART; SM00445; Link; 2.  
 PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 PROSITE; PS00615; C-TYPE\_LECTIN\_1; 1.  
 PROSITE; PS50041; C-TYPE\_LECTIN\_2; 1.  
 PROSITE; PS00022; EGF\_1; 2.  
 PROSITE; PS01186; EGF\_2; 1.  
 PROSITE; PS01187; EGF\_CA; 1.  
 PROSITE; PS01241; LINK; 2.  
 Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sush1;  
 Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
 Hyaluronic acid; Alternative splicing.  
 SIGNAL 1 20  
 CHAIN 21 3381  
 DOMAIN 37 138  
 FT DOMAIN 168 245  
 FT DOMAIN 266 347  
 FT DOMAIN 349 1336  
 FT GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT  
 DOMAIN).

FT	DOMAIN	1337	3074	GAG-BETA.	DE
FT	DOMAIN	3074	3110	EGF-LIKE 1.	GN
FT	DOMAIN	3112	3148	EGF-LIKE 2, CALCIUM-BINDING.	OS
FT	DOMAIN	3161	3275	C-TYPE LECTIN.	OC
FT	DOMAIN	3280	3338	SUSHI.	OC
FT	DISULFID	44	131	BY SIMILARITY.	OX
FT	DISULFID	173	244	BY SIMILARITY.	[1]
FT	DISULFID	197	218	BY SIMILARITY.	RP
FT	DISULFID	271	346	BY SIMILARITY.	RC
FT	DISULFID	295	316	BY SIMILARITY.	RX
FT	DISULFID	3078	3089	BY SIMILARITY.	RA
FT	DISULFID	3083	3098	BY SIMILARITY.	RA
FT	DISULFID	3100	3109	BY SIMILARITY.	RT
FT	DISULFID	3116	3127	BY SIMILARITY.	RL
FT	DISULFID	3121	3136	BY SIMILARITY.	RN
FT	DISULFID	3138	3147	BY SIMILARITY.	RP
FT	DISULFID	3154	3165	BY SIMILARITY.	RC
FT	DISULFID	3182	3274	BY SIMILARITY.	RX
FT	DISULFID	3250	3266	BY SIMILARITY.	RA
FT	DISULFID	3281	3324	BY SIMILARITY.	RA
FT	DISULFID	3310	3337	BY SIMILARITY.	RA
FT	CARBOHYD	57	57	N-LINKED (GLCNAC. . .)	RT
FT	CARBOHYD	331	331	N-LINKED (GLCNAC. . .)	RT
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. . .)	RT
FT	CARBOHYD	817	817	N-LINKED (GLCNAC. . .)	RL
FT	CARBOHYD	965	965	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1017	1017	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1333	1333	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1393	1393	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1437	1437	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1463	1463	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1653	1653	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1974	1974	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2045	2045	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2074	2074	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2103	2103	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2263	2263	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2290	2290	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2356	2356	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2623	2623	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2641	2641	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2919	2919	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	3052	3052	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	3354	3354	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	3364	3364	N-LINKED (GLCNAC. . .)	CC
FT	VARSPLIC	349	349	P -> R (IN ISOFORM V1 AND V3).	CC
FT	VARSPLIC	350	1336	MISSING (IN ISOFORM V1).	CC
FT	VARSPLIC	1337	3074	MISSING (IN ISOFORM V2).	CC
FT	CONFLICT	25	25	MISSING (IN ISOFORM V3).	CC
FT	CONFLICT	51	51	MISSING (IN REF. 2).	CC
FT	CONFLICT	89	89	N -> D (IN REF. 2).	CC
FT	CONFLICT	96	96	O -> D (IN REF. 2).	CC
FT	CONFLICT	346	346	C -> R (IN REF. 2).	DR
SQ	SEQUENCE	3381 AA;	369984 MW;	F09716FA7778D459 CRC64;	DR
Query Match				87.0%; Score 47; DB 1; Length 3381;	DR
Best Local Similarity				80.0%; Pred. No. 2.6;	DR
Matches				8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	DR
QY	1	CVIGXSGDRC 10			DR
Db	3100	CVPGYSGDRC 3109			DR
RESULT 4					DR
ID	NRG2_MOUSE	STANDARD;	PRT;	756 AA.	DR
AC	P56974;				DR
DT	16-OCT-2001 (Rel. 40, Created)				DR
DT	16-OCT-2001 (Rel. 40, Last sequence update)				DR
DT	16-OCT-2001 (Rel. 40, Last annotation update)				DR
DE	Pro-neuregulin-2 precursor (Pro-NRG2) [Contains: Neuregulin-2 (NRG-2)]				DR

(Divergent of neuregulin 1) (DON-1)).

NRG2.

Mus musculus (Mouse).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

SEQUENCE FROM N.A. (ISOFORMS NRG2-5; NRG2-10 AND NRG2-16A).

STRAIN=C57BL/6; TISSUE=Brain;

MEDLINE=97311398; PubMed=9168115;

Carraway K.L. III, Weber J.L., Unger M.J., Ledesma J., Yu N., Gassmann M., Lai C.;

"Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine kinases."

Nature 387:512-516(1997).

SEQUENCE OF 150-756 FROM N.A. (ISOFORMS DON-1M AND DON-1S).

TISSUE=Choroid plexus;

MEDLINE=97342638; PubMed=9199335;

Busfield S.J., Michnick D.A., Chickering T.W., Revett T.L., Ma J., Woolf E.A., Comrack C.A., Dussault B.J., Woolf J., Goodearl A.D.J., Gearing D.P.;

"Characterization of a neuregulin-related gene, Don-1, that is highly expressed in restricted regions of the cerebellum and hippocampus."

Mol. Cell. Biol. 17:4007-4014(1997).

-1- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS, RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND ACTIVATION OF THE ERBB RECEPTORS. MAY ALSO PROMOTE THE HETERODIMERIZATION WITH THE EGF RECEPTOR.

-1- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).

-1- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS: DON-1M, DON-1S/NRG2-5, NRG2-10 AND NRG2-16A (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.

-1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN THE BRAIN, WITH LOWER LEVELS IN THE LUNG. IN THE CEREBELLUM, FOUND IN GRANULE AND PURKINJE CELLS.

-1- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN DIMERIZATION (BY SIMILARITY).

-1- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE DOMAIN (BY SIMILARITY).

-1- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR FORM (BY SIMILARITY).

-1- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY SIMILARITY).

-1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.

-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.

-1- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.

HSSP: Q12784; IHR.

MGD; MGI:1098246; Nrg2.

InterPro: IPR000561; EGF-like.

InterPro: IPR003006; Ig\_MHC.

InterPro: IPR003598; Ig\_c2.

InterPro: IPR002154; Neuregulin.

Pfam: PF000008; EGF; 1.

Pfam: PF000047; Ig; 1.

Pfam: PF02158; Neuregulin; 1.

SMART: SM00181; EGF; 1.

SMART: SM00408; IGC2; 1.

PROSITE: PS00022; EGF\_1; 1.

PROSITE: PS01186; EGF\_2; 1.

Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein; Transmembrane; Multigene family; Alternative splicing.

PROPEP 1 19

CHAIN 20 756

CHAIN 20 314

DOMAIN 20 315

TRANSMEM 316 336

PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.

NEUREGULIN-2.

EXTRACELLULAR (POTENTIAL).

INTERNAL SIGNAL SEQUENCE (POTENTIAL).

```

FT DOMAIN 337 756 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 158 226 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 238 248 SER/THR-RICH.
FT DOMAIN 290 290 EGF-LIKE.
FT DOMAIN 627 633 POLY-PRO.
FT DISULFID 165 219 BY SIMILARITY.
FT DISULFID 253 267 BY SIMILARITY.
FT DISULFID 261 278 BY SIMILARITY.
FT DISULFID 280 289 BY SIMILARITY.
FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 C -> G (IN ISOFORM NRG2-10).
FT VARSPLIC 280 280 MISSING (IN ISOFORM NRG2-10).
FT VARSPLIC 281 756 VGYTDCRCQOFAMVNFSKHLGFKELKEAEELYOKRVLTITGI
FT VARSPLIC 282 330 CVALLIVG -> NFFGQRCLEKLPRLYMPDKQSVLWMDT
FT FT PGTGVSSQWSTSLDLN (IN ISOFORM DON-1S).
FT VARSPLIC 331 756 MISSING (IN ISOFORM DON-1S).
FT VARSPLIC 282 307 VGYTDCRCQOFAMVNFSKHLGFKELKE -> NGFFGQRCLEK
FT FT LPLRLYMPDKQK (IN ISOFORM DON-1M).
SQ SEQUENCE 756 AA; 82213 MW; 51D85DC918B678E CRC64;

Query Match 81.5%; Score 44; DB 1; Length 756;
Best Local Similarity 60.0%; Pred. No. 2.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
   I : : : : :
Db 280 CPVGYTGDCR 289

RESULT 5
EGF_RAT
ID EGF_RAT STANDARD; PRT; 1133 AA.
AC P07522; O63183;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal
DE growth factor].
DE EGF.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE-Kidney;
RC MEDLINE=92398779; PubMed=1524680;
RA Price P.M., Sagai S.J., Safirstein R.;
RT "Cloning and sequencing of the rat preproepidermal growth factor
RT cDNA: comparison with mouse and human sequences.";
RL DNA Cell Biol. 11:481-487(1992).
[2]
RN REVISIONS.
RP TISSUE-Kidney;
RC Price P.M.;
RA Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 974-1021.
RX MEDLINE=86081810; PubMed=3000782;
RA Simpson R.J., Smith J.A., Moritz R.L., O'Hare M.J., Rudland P.S.,
RA Morrison J.R., Lloyd C.J., Grego B., Burgess A.W., Nice E.C.;
RT "Rat epidermal growth factor: complete amino acid sequence. Homology
RT with the corresponding murine and human proteins; isolation of a form
RT truncated at both ends with full in vitro biological activity.";
RL Eur. J. Biochem. 153:629-637(1985).
[4]
RN SEQUENCE OF 994-1108 FROM N.A.
RP STRAIN-Sprague-Dawley; TISSUE-Kidney;
RX MEDLINE=89016634; PubMed=3262867;
RA Dorow D.S., Simpson R.J.;

```

```

RT "Cloning and sequence analysis of a cDNA for rat epidermal growth
RT factor.";
RL Nucleic Acids Res. 16:9338-9338(1988).
CC -|- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
CC FIBROBLASTS IN CELL CULTURE.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U04842; AAB60436.1; -.
DR EMBL; X12748; CAA31241.1; -.
DR PIR; A25425; EGR1.
DR HSSP; P01132; IEGF.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001336; EGF-1.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR000033; Ldl_receptor_rep.
DR Pfam; PF00008; EGF; 7.
DR Pfam; PF00058; ldl_recept_b; 7.
DR PRINTS; PR00009; EGFTGF.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF_like; 6.
DR SMART; SM00135; LY; 8.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 6.
DR PROSITE; PS01187; EGF_CA; 3.
KW EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;
KW Signal.
FT CHAIN 1 21 POTENTIAL.
FT CHAIN 22 1133 PRO-EPIDERMAL GROWTH FACTOR.
FT DOMAIN 974 1026 EPIDERMAL GROWTH FACTOR.
FT TRANSMEM 1036 1057 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1058 1133 POTENTIAL.
FT DOMAIN 322 356 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 357 397 EGF-LIKE 1 (INCOMPLETE).
FT DOMAIN 398 438 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 436 478 EGF-LIKE 3.
FT DOMAIN 743 783 EGF-LIKE 4.
FT DOMAIN 835 873 EGF-LIKE 5.
FT DOMAIN 874 915 EGF-LIKE 6.
FT DOMAIN 916 956 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 975 1016 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DISULFID 361 372 EGF-LIKE 9.
FT DISULFID 368 381 BY SIMILARITY.
FT DISULFID 383 396 BY SIMILARITY.
FT DISULFID 402 413 BY SIMILARITY.
FT DISULFID 409 422 BY SIMILARITY.
FT DISULFID 424 437 BY SIMILARITY.
FT DISULFID 440 452 BY SIMILARITY.
FT DISULFID 448 462 BY SIMILARITY.
FT DISULFID 464 477 BY SIMILARITY.
FT DISULFID 747 758 BY SIMILARITY.
FT DISULFID 754 767 BY SIMILARITY.
FT DISULFID 769 782 BY SIMILARITY.
FT DISULFID 839 850 BY SIMILARITY.
FT DISULFID 844 859 BY SIMILARITY.
FT DISULFID 861 872 BY SIMILARITY.
FT DISULFID 878 892 BY SIMILARITY.
FT DISULFID 885 901 BY SIMILARITY.
FT DISULFID 903 914 BY SIMILARITY.
FT DISULFID 920 933 BY SIMILARITY.
FT DISULFID 927 942 BY SIMILARITY.

```

```

FT DISULFID 944 955 BY SIMILARITY.
FT DISULFID 979 993
FT DISULFID 987 1004
FT DISULFID 1006 1015
FT CARBOHYD 105 108
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 930 930 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 941 941 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 955 955 C -> V.
FT VARIANT 1024 1025 KL -> NW (IN REF. 4).
FT CONFLICT 1108 1108 Q -> S (IN REF. 4).
SQ SEQUENCE 1133 AA; 124125 MW; C224A302E9578031 CRC64;

Query Match 81.5%; Score 44; DB 1; Length 1133;
Best Local Similarity 70.0%; Pred. No. 3.1; Indels 0; Gaps 0;
Matches 7; Conservative 2; Mismatches 1;

QY 1 CVIGXSGDRC 10
Db 1006 CVIGIGERC 1015
||||:|:|

RESULT 6
PGCV_MACNE STANDARD; PRT; 862 AA.
AC Q28858; Q28859; Q28860;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Versican core protein (large fibroblast proteoglycan) (Chondroitin
DE sulfate proteoglycan core protein 2) (Fragments).
GN CSPG2.
OS Macaca nemestrina (Pig-tailed macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9545;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Aortic smooth muscle;
RA MEDLINE=9500762; PubMed=7921538;
RX Yao L.Y., Moody C., Schoenherr E., Wight T.N., Sandell L.J.;
RT "Identification of the proteoglycan versican in aorta and smooth
RT muscle cells by DNA sequence analysis, in situ hybridization and
RT immunohistochemistry.";
RL Matrix Biol. 14:213-225(1994).
CC -|- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronan.
CC -|- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC similarity).
CC -|- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (By similarity).
CC -|- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -|- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -|- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; S72412; AAA65593.2; -
CC EMBL; S72413; AAA65594.2; -
CC EMBL; S72414; AAA65595.2; -
CC HSP; P01132; IEGF.

```

```

DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR PRINTS; PR00010; EGFBL00D.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00445; LINK; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; PARTIAL.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Repeat;
KW EGF-like domain; Calcium.
FT NON_TER 1 1
FT DOMAIN <1 37 LINK 1.
FT DOMAIN 58 139 LINK 2.
FT DOMAIN 141 >233 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
FT DOMAIN) (DOMAIN).
FT NON_CONS 233 234
FT DOMAIN <234 >525 GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN,
FT SIMILAR TO CHONDROITIN SULFATE ATTACHMENT
FT SITE IN COLLAGEN TYPE IX (BY SIMILARITY).
FT NON_CONS 525 526 EGF-LIKE 1.
FT DOMAIN 718 754 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 756 792 C-TYPE LECTIN.
FT DOMAIN 805 >862 BY SIMILARITY.
FT DISULFID 63 138 BY SIMILARITY.
FT DISULFID 87 108 BY SIMILARITY.
FT DISULFID 722 733 BY SIMILARITY.
FT DISULFID 727 742 BY SIMILARITY.
FT DISULFID 744 753 BY SIMILARITY.
FT DISULFID 760 771 BY SIMILARITY.
FT DISULFID 765 780 BY SIMILARITY.
FT DISULFID 782 791 BY SIMILARITY.
FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 563 563 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 696 696 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 862 862
SQ SEQUENCE 862 AA; 95583 MW; A5D5F6153A74BB39 CRC64;

Query Match 79.6%; Score 43; DB 1; Length 862;
Best Local Similarity 70.0%; Pred. No. 3.6;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 744 CVPYSGDQC 753
||||:|:|

RESULT 7
EGF_HUMAN STANDARD; PRT; 1207 AA.
AC P01133;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal
DE growth factor (Urogastrone)].
GN EGF.

```



CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RX [1]  
RP SEQUENCE FROM N.A. (ISOFORM V0).  
RC MEDLINE=95105188; PubMed=7528742;  
RX Naso M.F., Zimmermann D.R., Iozzo R.V.;  
RA "Characterization of the complete genomic structure of the human  
RT versican gene and functional analysis of its promoter.";  
RL J. Biol. Chem. 269:32999-33008(1994).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM V1).  
RC TISSUE=Placenta;  
RX MEDLINE=90059882; PubMed=2593089;  
RA Zimmermann D.R., Ruoslahti E.;  
RT "Multiple domains of the large fibroblast proteoglycan, versican.";  
RL EMBO J. 8:2975-2981(1989).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM V2).  
RC TISSUE=Glial tumor;  
RX MEDLINE=95105187; PubMed=7806529;  
RA Dours-Zimmermann M.T., Zimmermann D.R.;  
RT "A novel glycosaminoglycan attachment domain identified in two  
RT alternative splice variants of human versican.";  
RL J. Biol. Chem. 269:32992-32998(1994).  
RN [4]  
RP SEQUENCE OF 2711-3396 FROM N.A.  
RC TISSUE=Lung fibroblast;  
RX MEDLINE=88007514; PubMed=2820964;  
RA Krusius T., Gehlsen K.R., Ruoslahti E.;  
RT "A fibroblast chondroitin sulfate proteoglycan core protein contains  
RT lectin-like and growth factor-like sequences.";  
RL J. Biol. Chem. 262:13120-13125(1987).  
RN [5]  
RP SEQUENCE OF 251-347 FROM N.A.  
RX MEDLINE=93122792; PubMed=1478664;  
RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,  
RA McPherson J.D.;  
RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of  
RT human chromosome 5 (5q12-5q14).";  
RL Genomics 14:845-851(1992).  
RN [6]  
RP SEQUENCE FROM N.A. (ISOFORM V3).  
RC TISSUE=Brain;  
RX MEDLINE=95181355; PubMed=7876137;  
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;  
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M  
RT without a chondroitin sulfate attachment in region in mouse and human  
RT tissues.";  
RL J. Biol. Chem. 270:3914-3918(1995).  
RN [7]  
RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).  
RC TISSUE=Aortic smooth muscle;  
RX MEDLINE=99327053; PubMed=10397680;  
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,  
RA Wight T.N.;  
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";  
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).  
RN [8]  
RP PARTIAL SEQUENCE.  
RC TISSUE=Brain;  
RX MEDLINE=89174663; PubMed=2466833;  
RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;  
RT "Isolation and partial characterization of a glial  
RT hyaluronate-binding protein.";  
RL J. Biol. Chem. 264:5981-5987(1989).  
RN [9]  
RP TISSUE SPECIFICITY OF ISOFORMS.  
RX MEDLINE=96213482; PubMed=8627343;  
RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;  
RT "Differential expression of versican isoforms in brain tumors.";  
RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).  
CC -!- FUNCTION: May play a role in intercellular signaling and in  
connecting cells with the extracellular matrix. May take part in

CC the regulation of cell motility, growth and differentiation. Binds  
CC hyaluronic acid.  
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
CC -!- ALTERNATIVE PRODUCTS: At least 5 isoforms: V0 (shown here), V1,  
CC V2, V3 and Vint; are produced by alternative splicing.  
CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed  
CC in normal brain, gliomas, medulloblastomas, schwannomas,  
CC neurofibromas, and meningiomas; V2 is restricted to normal brain  
CC and gliomas; V3 is found in all these tissues except  
CC medulloblastomas.  
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.  
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL: U16306; AAA65018.1; -  
CC EMBL: X15998; CAA34128.1; -  
CC EMBL: S52488; AAB24878.1; -  
CC EMBL: U28555; AAA67565.1; -  
CC EMBL: J32039; BAA06801.1; -  
CC EMBL: J02814; AAA36437.1; -  
CC EMBL: AF084545; AAD48545.1; -  
CC PIR: S06014; S06014.  
CC PIR: A29348; A29348.  
CC PIR: A30358; A30358.  
CC HSP: P01132; LEFG.  
CC Genew: HGNC:2464; CSPG2.  
CC MIM: 118661; -  
CC InterPro: IPR000152; Asx\_hydroxyl.  
CC InterPro: IPR000561; EGF-like.  
CC InterPro: IPR000742; EGF\_2.  
CC InterPro: IPR001881; EGF\_Ca.  
CC InterPro: IPR001438; EGF\_II.  
CC InterPro: IPR003599; Ig.  
CC InterPro: IPR003006; Ig\_MHC.  
CC InterPro: IPR001304; Lectin\_C.  
CC InterPro: IPR000538; Link.  
CC InterPro: IPR000436; Sushi\_SCR\_CCP.  
CC Pfam: PF00008; EGF; 2.  
CC Pfam: PF00047; Ig; 1.  
CC Pfam: PF00059; Lectin\_c; 1.  
CC Pfam: PF00084; sushi; 1.  
CC Pfam: PF00193; Xlink; 2.  
CC PRINTS: PR00010; EGF\_BLOOD.  
CC PRODOM: PD000918; Link; 2.  
CC SMART: SM00032; CCP; 1.  
CC SMART: SM00034; CLECT; 1.  
CC SMART: SM00179; EGF\_Ca; 1.  
CC SMART: SM00001; EGF\_like; 1.  
CC SMART: SM00409; IG; 1.  
CC SMART: SM00445; LINK; 2.  
CC PROSITE: PS00010; ASX\_HYDROXYL; 1.  
CC PROSITE: PS00022; EGF\_1; 2.  
CC PROSITE: PS01186; EGF\_2; 1.  
CC PROSITE: PS01187; EGF\_Ca; 1.  
CC PROSITE: PS01241; LINK; 2.  
CC PROSITE: PS00615; C-TYPE LECTIN; 1.  
CC PROSITE: PS00041; C-TYPE LECTIN; 1.  
CC Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;  
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
FT Hyaluronic acid; Alternative splicing.  
FT SIGNAL 1 20 POTENTIAL.





```
FT DISULFID 476 485 BY SIMILARITY.
FT DISULFID 492 503 BY SIMILARITY.
FT DISULFID 497 512 BY SIMILARITY.
FT DISULFID 514 523 BY SIMILARITY.
FT DISULFID 530 541 BY SIMILARITY.
FT DISULFID 535 550 BY SIMILARITY.
FT DISULFID 552 561 BY SIMILARITY.
FT DISULFID 636 647 BY SIMILARITY.
FT DISULFID 641 656 BY SIMILARITY.
FT DISULFID 658 667 BY SIMILARITY.
FT DISULFID 674 685 BY SIMILARITY.
FT DISULFID 679 694 BY SIMILARITY.
FT DISULFID 696 705 BY SIMILARITY.
FT DISULFID 712 723 BY SIMILARITY.
FT DISULFID 717 732 BY SIMILARITY.
FT DISULFID 734 743 BY SIMILARITY.
FT DISULFID 751 762 BY SIMILARITY.
FT DISULFID 756 771 BY SIMILARITY.
FT DISULFID 773 782 BY SIMILARITY.
FT DISULFID 789 800 BY SIMILARITY.
FT DISULFID 794 809 BY SIMILARITY.
FT DISULFID 811 820 BY SIMILARITY.
FT DISULFID 827 838 BY SIMILARITY.
FT DISULFID 832 847 BY SIMILARITY.
FT DISULFID 849 858 BY SIMILARITY.
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 385 385 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1046 1046 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1242 AA; 135343 MW; 59557A4B7F052860 CRC64;

Query Match 77.8%; Score 42; DB 1; Length 1242;
Best Local Similarity 60.0%; Pred. No. 8;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
   | 1:|||||
DB 476 CAAGFAGDRC 485

RESULT 10
FBP1_STRPU STANDARD; PRT; 1064 AA.
AC P10079;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fibropellin I precursor (Epidermal growth factor-related protein 1)
DE (UEGF-1).
GN EGF1.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_Taxid=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90112459; PubMed=2514273;
RA Delgadillo-Reynoso M.G., Rollo D.R., Hursh D.A., Raff R.A.;
RT "Structural analysis of the uegf gene in the sea urchin
RT strongylocentrotus purpuratus reveals more similarity to vertebrate
RT than to invertebrate genes with EGF-like repeats.";
RL J. Mol. Evol. 29:314-327(1989).
RN [2]
RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX MEDLINE=87319677; PubMed=3498216;
RA Hursh D.A., Andrews M.E., Raff R.A.;
RT "A sea urchin gene encodes a polypeptide homologous to epidermal
RT growth factor.";
```

```
RL Science 237:1487-1490(1987).
RN [3]
RP AVIDIN-LIKE DOMAIN.
RX MEDLINE=89196806; PubMed=2784773;
RA Hunt L.T., Barker W.C.;
RT "Avidin-like domain in an epidermal growth factor homolog from a sea
RT urchin.";
RL FASEB J. 3:1760-1764(1989).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=91285254; PubMed=2060714;
RA Bisgrove B.W., Andrews M.E., Raff R.A.;
RT "Fibropellins, products of an EGF repeat-containing gene, form a
RT unique extracellular matrix structure that surrounds the sea urchin
RT embryo.";
RL Dev. Biol. 146:89-99(1991).
CC -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC MATRIX.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC EMBRYOS AND EARLY LARVAE.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1A (SHOWN HERE) AND 1B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (1B) LACKS 8 EGF
CC REPEATS.
CC -!- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC AND ZYGOTICALLY.
CC -!- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC TO AVIDIN/STREPTAVIDIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L08692; AA62164.1; -
DR EMBL; L08692; AA62163.1; -
DR EMBL; X17530; CAA35571.1; -
DR EMBL; M17421; AAA30050.1; -
DR EMBL; X17533; CAA35573.1; -
DR PIR; A29316; A29316.
DR HSP; P01132; 1EGF.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000088; Avidin.
DR InterPro; IPR000089; CUB.domain.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00008; EGF; 21.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF01382; Avidin; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00042; CUB; 1.
DR SMART; SM00179; EGF_CA; 18.
DR SMART; SM00001; EGF-like; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 19.
DR PROSITE; PS00022; EGF_1; 19.
DR PROSITE; PS00577; AVIDIN; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01186; EGF_2; 19.
DR PROSITE; PS01187; EGF_CA; 18.
KW Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;
```

KW Glycoprotein; Calcium-binding.  
FT SIGNAL 1 19  
FT CHAIN 20 1064 POTENTIAL.  
FT DOMAIN 20 55 FIBROPELIN I.  
FT DOMAIN 62 175 EGF-LIKE 1.  
FT DOMAIN 176 212 CUB.  
FT DOMAIN 212 250 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 366 402 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 442 478 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 480 516 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 518 554 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 556 592 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 594 630 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 632 668 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 670 706 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 708 744 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 746 782 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 784 820 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 822 858 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 860 896 EGF-LIKE 20.  
FT DOMAIN 898 934 AVIDIN-LIKE.  
FT DOMAIN 936 1064 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).  
FT DISULFID 23 34 BY SIMILARITY.  
FT DISULFID 28 43 BY SIMILARITY.  
FT DISULFID 45 54 BY SIMILARITY.  
FT DISULFID 180 191 BY SIMILARITY.  
FT DISULFID 185 200 BY SIMILARITY.  
FT DISULFID 202 211 BY SIMILARITY.  
FT DISULFID 218 229 BY SIMILARITY.  
FT DISULFID 223 238 BY SIMILARITY.  
FT DISULFID 240 249 BY SIMILARITY.  
FT DISULFID 256 267 BY SIMILARITY.  
FT DISULFID 261 276 BY SIMILARITY.  
FT DISULFID 278 305 BY SIMILARITY.  
FT DISULFID 294 305 BY SIMILARITY.  
FT DISULFID 299 314 BY SIMILARITY.  
FT DISULFID 316 325 BY SIMILARITY.  
FT DISULFID 332 343 BY SIMILARITY.  
FT DISULFID 337 352 BY SIMILARITY.  
FT DISULFID 354 363 BY SIMILARITY.  
FT DISULFID 370 381 BY SIMILARITY.  
FT DISULFID 375 390 BY SIMILARITY.  
FT DISULFID 392 401 BY SIMILARITY.  
FT DISULFID 408 419 BY SIMILARITY.  
FT DISULFID 413 428 BY SIMILARITY.  
FT DISULFID 430 439 BY SIMILARITY.  
FT DISULFID 446 457 BY SIMILARITY.  
FT DISULFID 451 466 BY SIMILARITY.  
FT DISULFID 468 477 BY SIMILARITY.  
FT DISULFID 484 495 BY SIMILARITY.  
FT DISULFID 489 504 BY SIMILARITY.  
FT DISULFID 506 515 BY SIMILARITY.  
FT DISULFID 522 533 BY SIMILARITY.  
FT DISULFID 527 542 BY SIMILARITY.  
FT DISULFID 544 553 BY SIMILARITY.  
FT DISULFID 560 571 BY SIMILARITY.  
FT DISULFID 565 580 BY SIMILARITY.  
FT DISULFID 582 591 BY SIMILARITY.  
FT DISULFID 598 609 BY SIMILARITY.  
FT DISULFID 603 618 BY SIMILARITY.  
FT DISULFID 620 629 BY SIMILARITY.  
FT DISULFID 636 647 BY SIMILARITY.  
FT DISULFID 641 656 BY SIMILARITY.  
FT DISULFID 658 667 BY SIMILARITY.  
FT DISULFID 674 685 BY SIMILARITY.  
FT DISULFID 679 694 BY SIMILARITY.  
FT DISULFID 696 705 BY SIMILARITY.  
FT DISULFID 712 723 BY SIMILARITY.  
FT DISULFID 717 732 BY SIMILARITY.  
FT DISULFID 734 743 BY SIMILARITY.  
FT DISULFID 750 761 BY SIMILARITY.  
FT DISULFID 755 770 BY SIMILARITY.  
FT DISULFID 772 781 BY SIMILARITY.  
FT DISULFID 788 799 BY SIMILARITY.  
FT DISULFID 793 808 BY SIMILARITY.  
FT DISULFID 810 819 BY SIMILARITY.  
FT DISULFID 826 837 BY SIMILARITY.  
FT DISULFID 831 846 BY SIMILARITY.  
FT DISULFID 848 857 BY SIMILARITY.  
FT DISULFID 864 875 BY SIMILARITY.  
FT DISULFID 869 884 BY SIMILARITY.  
FT DISULFID 886 895 BY SIMILARITY.  
FT DISULFID 902 913 BY SIMILARITY.  
FT DISULFID 907 922 BY SIMILARITY.  
FT DISULFID 924 933 BY SIMILARITY.  
FT CARBOHYD 30 30 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 136 136 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 851 851 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT VARSPLIC 477 780 MISSING (IN ISOFORM IB).  
FT CONFLICT 279 279 L -> S (IN REF. 2).  
SQ SEQUENCE 1064 AA; 112072 MW; 2E569CA012ED6D09 CRC64;  
Query Match 75.9%; Score 41; DB 1; Length 1064;  
Best Local Similarity 60.0%; Pred. NO. 10;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CVIGXSGDRC 10  
DB 316 CPLGFSGDC 325  
RESULT 11  
CLR2\_MOUSE STANDARD; PRT: 2920 AA.  
ID CLR2\_MOUSE Q9Z2R4; Q99K26;  
AC Q9ROM0: Q9Z2R4; Q99K26;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Flamingo 1)  
DE (mFmil).  
GN CELSR2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99418630; PubMed=10490098;  
RA Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,  
RA Takeichi M., Uemura T.;  
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell  
RL polarity under the control of frizzled.";  
RN Cell 98:585-595(1999).  
RP [2]  
RP SEQUENCE OF 1913-2796 FROM N.A., AND TISSUE SPECIFICITY.  
RX PubMed=10790539;  
RA Formstone C.J., Barclay J., Rees M., Little P.F.R.;  
RT "Chromosomal localization of Celser2 and Celser3 in the mouse; Celser3 is  
RL a candidate for the tippy (tip) lethal mutant on chromosome 9.";  
RN Mamm. Genome 11:392-394(2000).  
RP [3]  
RP SEQUENCE OF 2014-2920 FROM N.A.  
RC TISSUE-Breast tumor;  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
RP [4]  
RP DEVELOPMENTAL STAGE.  
RX PubMed=11850187;  
RA Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;  
RT "Developmental expression profiles of Celser (Flamingo) genes in the  
RL mouse.";  
RL Mech. Dev. 112:157-160(2002).



```
FT CARBOHYD 1502 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1742 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1828 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2025 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2044 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2062 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2346 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 2199 V -> L (IN REF. 3).
FT CONFLICT 2283 A -> V (IN REF. 3).
FT CONFLICT 2535 S -> R (IN REF. 1).
FT CONFLICT 2571 L -> R (IN REF. 2).
FT CONFLICT 2639 Y -> S (IN REF. 3).

Query Match 75.9% Score 41; DB 1; Length 2920;
Best Local Similarity 60.0%; Pred. No. 29;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 1820 CVLGYGDC 1829

RESULT 12
EGF_PIG STANDARD; PRT; 53 AA.
AC Q00968;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Epidermal growth factor (EGF) (Fragment).
GN EGF.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9623;
RN [1]
RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX TISSUE=Kidney;
RA Pascall J.C., Jones D.S.C., Doel S.M., Clements J.M., Hunter M.,
RA Fallon T., Edwards M., Brown K.D.;
RT "Cloning and characterization of a gene encoding pig epidermal growth
RT factor.";
RL J. Mol. Endocrinol. 6:63-70(1991).
CC -!- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
CC FIBROBLASTS IN CELL CULTURE.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59516; CAA42102.1; -
DR HSP; P01132; LEGF
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
KW EGF-like domain; Growth factor.
FT NON_TER 1 1
FT PEPTIDE 1 53 EPIDERMAL GROWTH FACTOR.
FT DOMAIN 2 43 EGF-LIKE.
FT DISULFID 6 20 BY SIMILARITY.
FT DISULFID 14 31 BY SIMILARITY.
```

```
FT DISULFID 33 42 BY SIMILARITY.
FT NON_TER 53 53
SQ SEQUENCE 53 AA; 6149 MW; 74F615B4A05774D4 CRC64;

Query Match 74.1%; Score 40; DB 1; Length 53;
Best Local Similarity 60.0%; Pred. No. 0.79;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 33 CVFGYVGERC 42

RESULT 13
Y402_ARCFU STANDARD; PRT; 57 AA.
ID Y402_ARCFU STANDARD; PRT; 57 AA.
AC O29845;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF0402.
GN AF0402.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF001076; AAB90836.1; -
DR TIGR; AF0402; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 57 AA; 6885 MW; 4C71BDEB722568B8 CRC64;

Query Match 74.1%; Score 40; DB 1; Length 57;
Best Local Similarity 50.0%; Pred. No. 0.85;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 23 CIMGYTGKRC 32

RESULT 14
NTC2_MOUSE STANDARD; PRT; 2470 AA.
ID NTC2_MOUSE STANDARD; PRT; 2470 AA.
AC O35516; Q06008;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
```

15-JUN-2002 (Rel. 41, Last annotation update)  
Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch  
B).  
NOTCH2.  
Mus musculus (Mouse).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=C57BL/6; TISSUE=Thymus;  
Hamada Y., Higuchi M., Tsujimoto Y.;  
RT "Complete amino acid sequence and multiform transcripts encoded by a  
RT single copy of mouse Notch2 gene.";  
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
[2]  
SEQUENCE OF 316-1518 FROM N.A.  
STRAIN=C57BL/6 X CBA; TISSUE=Embryo;  
MEDLINE=93178563; PubMed=8440332;  
RA Lardelli M., Lendahl U.;  
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a  
RT wide variety of tissues.";  
RL Exp. Cell Res. 204:364-372(1993).  
[3]  
SEQUENCE OF 1765-2153 FROM N.A.  
MEDLINE=97075110; PubMed=8917536;  
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,  
RA Martin D.I.;  
RT "Inhibition of granulocytic differentiation by mNotch1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).  
[4]  
FUNCTION.  
RP MEDLINE=99396706; PubMed=10393120;  
RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,  
RA Tsujimoto Y.;  
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early  
RT embryonic lethality.";  
RL Development 126:3415-3424(1999).  
[5]  
DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.  
RP MEDLINE=95333893; PubMed=7609614;  
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;  
RT "Differential expression of Notch1 and Notch2 in developing and adult  
RT mouse brain.";  
RL Brain Res. Mol. Brain Res. 29:263-272(1995).  
[6]  
POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
RP MEDLINE=21523956; PubMed=11518718;  
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;  
RT "Murine notch homologs (N1-4) undergo presenilin-dependent  
RT proteolysis.";  
RL J. Biol. Chem. 276:40268-40273(2001).  
[7]  
POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
RP MEDLINE=21374376; PubMed=11459941;  
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;  
RT "Conservation of the biochemical mechanisms of signal transduction  
RT among mammalian Notch family members.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).  
[8]  
FUNCTION. Functions as a receptor for membrane-bound ligands  
CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.  
CC Upon ligand activation through the released notch intracellular  
CC domain (NICD) it forms a transcriptional activator complex with  
CC RBP-J kappa and activates genes of the enhancer of split locus.  
CC Affects the implementation of differentiation, proliferation and  
CC apoptotic programs (by similarity). May play an essential role in  
CC postimplantation development, probably in some aspect of cell  
CC specification and/or differentiation.  
-1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
CC terminal fragment N(EC) which are probably linked by disulfide  
CC bonds.  
-1- SUBCELLULAR LOCATION: Type I membrane protein. Following  
CC proteolytical processing NICD is translocated to the nucleus.

-1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
produced by alternative splicing.  
-1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,  
neuroepithelia, somites, optic vesicles and branchial arches, but  
not heart.  
-1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,  
the postnatal ependymal cells, and the choroid plexus throughout  
embryonic and postnatal development.  
-1- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
which is proteolytically cleaved by a furin-like convertase in the  
trans-Golgi network before it reaches the plasma membrane to yield  
an active, ligand-accessible form. Cleavage results in a C-  
terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
ligand binding, it is cleaved by TNF-alpha converting enzyme  
(TACE) to yield a membrane-associated intermediate fragment called  
notch extracellular truncation (NEXT). This fragment is then  
cleaved by presenilin dependent gamma-secretase to release a  
notch-derived peptide containing the intracellular domain (NICD)  
from the membrane.  
-1- PTM: Phosphorylated.  
-1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.  
-1- SIMILARITY: CONTAINS 34, 5 EGF-LIKE DOMAINS.  
-1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.  
-1- SIMILARITY: CONTAINS 6 ANK REPEATS.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; D32210; BAA22094.1; -  
EMBL; X68279; CAA48340.1; -  
EMBL; U31881; AAC52924.1; -  
HSSP; P16109; IFSB.  
DR MGD; MGI:97364; Notch2.  
DR InterPro; IPR002110; ANK.  
DR InterPro; IPR000152; ASX\_hydroxyl.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR000742; EGF-2.  
DR InterPro; IPR001881; EGF\_CA.  
DR InterPro; IPR001438; EGF-II.  
DR InterPro; IPR000800; Notch.  
DR Pfam; PF00008; EGF; 35.  
DR Pfam; PF00023; ank; 6.  
DR Pfam; PF00066; notch; 2.  
DR PRINTS; PR00010; EGFBLD.  
DR PRINTS; PR01452; NOTCH.  
DR SMART; SM00248; ANK; 4.  
DR SMART; SM00179; EGF\_CA; 22.  
DR SMART; SM00001; EGF\_like; 12.  
DR SMART; SM00004; NL; 3.  
DR PROSITE; PS50088; ANK\_REPEAT; 4.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 22.  
DR PROSITE; PS00022; EGF\_1; 33.  
DR PROSITE; PS01186; EGF\_2; 27.  
DR PROSITE; PS01187; EGF\_CA; 22.  
DR PROSITE; PS01188; EGF\_CA; 22.  
KW Receptor; Transcription regulation; Activator; Differentiation;  
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;  
KW Transmembrane; Glycoprotein; Signal; Phosphorylation;  
KW Alternative splicing.  
FT SIGNAL 1 25  
FT CHAIN 26 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.  
FT CHAIN 1566 2470 NOTCH EXTRACELLULAR TRUNCATION.  
FT CHAIN 1697 2470 NOTCH INTRACELLULAR DOMAIN.  
FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1678 1698 POTENTIAL.  
FT DOMAIN 1699 2470 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 26 63 EGF-LIKE 1.  
FT DOMAIN 64 102 EGF-LIKE 2.









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 41.3333 Seconds  
(without alignments)  
49.850 Million cell updates/sec

Title: US-09-673-785D-8

Perfect score: 54

Sequence: 1 CVIGXSGDRC 10

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rhodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	1200	11 Q8VD07	Q8vd07 mus musculus
2	49	90.7	2146	5 Q9VC97	Q9vc97 drosophila
3	44	81.5	106	4 Q9UN93	Q9un93 homo sapien
4	44	81.5	308	4 Q9UN94	Q9un94 homo sapien
5	44	81.5	366	4 Q9UN95	Q9un95 homo sapien
6	44	81.5	372	5 Q21756	Q21756 caenorhabdi
7	44	81.5	877	4 Q9H306	Q9h3q6 homo sapien
8	44	81.5	878	4 Q9GZ22	Q9gzz2 homo sapien
9	44	81.5	878	4 Q9H3Q7	Q9h3q7 homo sapien
10	44	81.5	901	4 Q9H195	Q9h195 homo sapien
11	44	81.5	957	4 Q14651	Q14651 homo sapien
12	44	81.5	1217	4 Q9UKW9	Q9ukw9 homo sapien
13	43	79.6	89	6 Q28867	Q28867 equus cabal
14	43	79.6	264	5 Q20043	Q20043 caenorhabdi
15	42	77.8	165	13 Q90XX0	Q90xx0 oreocolobus
16	42	77.8	1162	5 Q8WTP0	Q8wtp0 halocynthia

17	42	77.8	2352	5	Q61240	Q61240 halocynthia
18	42	77.8	2531	5	Q16004	Q16004 lytechinus
19	42	77.8	4006	11	Q35452	Q35452 mus musculus
20	42	77.8	4114	11	Q54796	Q54796 mus musculus
21	41	75.9	379	11	Q35883	Q35883 rattus norv
22	41	75.9	447	11	Q63348	Q63348 rattus norv
23	41	75.9	601	10	Q9LYF1	Q9lyf1 arabidopsis
24	41	75.9	1531	11	Q9WVB5	Q9wvb5 mus musculus
25	41	75.9	2920	11	Q9R0M0	Q9r0m0 mus musculus
26	41	75.9	2972	5	Q90891	Q90891 caenorhabdi
27	40	74.1	601	5	Q22913	Q22913 caenorhabdi
28	40	74.1	708	5	Q9NFS9	Q9nfs9 drosophila
29	40	74.1	969	4	Q96KG6	Q96kg6 homo sapien
30	40	74.1	1035	5	Q9NEG1	Q9neg1 drosophila
31	40	74.1	1210	6	Q95ND4	Q95nd4 felis silve
32	40	74.1	1214	6	Q9BDQ0	Q9bdq0 sus scrofa
33	40	74.1	1216	6	Q9BEA0	Q9bea0 canis famil
34	40	74.1	2524	5	Q9GPA5	Q9gpa5 branchiosto
35	40	74.1	4072	5	Q9W4Y4	Q9w4y4 drosophila
36	39	72.2	191	5	Q8T5Z1	Q8t5z1 caenorhabdi
37	39	72.2	191	5	Q8T5Z0	Q8t5z0 caenorhabdi
38	39	72.2	193	5	Q8T5Z2	Q8t5z2 caenorhabdi
39	39	72.2	193	5	Q8STG0	Q8stg0 caenorhabdi
40	39	72.2	194	5	Q8T5Z3	Q8t5z3 caenorhabdi
41	39	72.2	473	4	Q9BW29	Q9bw29 homo sapien
42	39	72.2	473	4	Q12891	Q12891 homo sapien
43	39	72.2	615	13	Q57409	Q57409 brachydanio
44	39	72.2	963	5	Q9GPM9	Q9gpm9 caenorhabdi
45	39	72.2	1045	5	Q8T3A6	Q8t3a6 caenorhabdi

#### ALIGNMENTS

RESULT 1  
Q8VD07 PRELIMINARY; PRT; 1200 AA.  
AC Q8VD07;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Similar to epidermal growth factor.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC017681; AAHL7681.1; -;  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001336; EGF\_1.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR000033; Ldl\_receptor\_rep.  
DR Pfam; PF00008; EGF; 8.  
DR Pfam; PF00058; ldl\_recept\_b; 7.  
DR PRINTS; PRO0009; EGFGF.  
DR SMART; SM00181; EGF; 9.  
DR SMART; SM00179; EGF\_CA; 8.  
DR SMART; SM00135; LV; 9.  
DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_3.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
DR PROSITE; PS01186; EGF\_2; UNKNOWN\_6.  
DR PROSITE; PS01187; EGF\_CA; UNKNOWN\_3.  
SQ SEQUENCE 1200 AA; 131317 MW; B6A44F7294746476 CRC64;

Query Match 100.0%; Score 54; DB 11; Length 1200;

Best Local Similarity 90.0%; Pred. No. 0.13; Indels 0; Gaps 0;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CVIGXSGDRC 10
Db 992 CVIGSGDRC 1001

RESULT 2
Q9VC97 Q9VC97 PRELIMINARY; PRT; 2146 AA.
AC Q9VC97;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CRB protein.
GN CRB OR CG6383.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2183-2195(2000).
DR EMBL; AF003747; AAF56276.1; -.
DR HSSP; P00740; 1EDM.
DR FlyBase; FBgn0000368; crb.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF000008; EGF; 26.
DR Pfam; PF00054; laminin_G; 3.

```

---

```

DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGFBLAMININ.
DR SMART; SM00179; EGF_CA; 11.
DR SMART; SM00001; EGF_like; 16.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 16.
DR PROSITE; PS00022; EGF_1; 25.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS01187; EGF_CA; 13.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 2146 AA; 233570 MW; 8E2B9B32B761115 CRC64;

Query Match 90.7%; Score 49; DB 5; Length 2146;
Best Local Similarity 70.0%; Pred. NO. 2.2;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDRC 10
Db 569 CAVIGSGDRC 578
I : : : : :
I : : : : :

RESULT 3
Q9UN93 Q9UN93 PRELIMINARY; PRT; 106 AA.
AC Q9UN93;
ID Q9UN93;
AD Q9UN93;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Mucin 3 (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=99333563; PubMed=10405327;
RA Williams S.J., Munster D.J., Quin R.J., Gotley D.C., McGuckin M.A.;
RT "The MUC3 gene encodes a transmembrane mucin and is alternatively
RT spliced."
RL Biochem. Biophys. Res. Commun. 261:83-89(1999).
DR EMBL; AF143373; AAD45884.1; -.
DR InterPro; IPR000561; EGF-like.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 106 AA; 11340 MW; 5A2E3C24C905E182 CRC64;

Query Match 81.5%; Score 44; DB 4; Length 106;
Best Local Similarity 70.0%; Pred. NO. 0.88;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDRC 10
Db 14 CLPGFSGDRC 23
I : : : : :
I : : : : :

RESULT 4
Q9UN94 Q9UN94 PRELIMINARY; PRT; 308 AA.
AC Q9UN94;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Mucin 3 (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```



```

DR EMBL; AB038784; BAB12118.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER 1
SQ SEQUENCE 877 AA; 93720 MW; 09EE5B86BC83BDFA CRC64;

Query Match      81.5%; Score 44; DB 4; Length 877;
Best Local Similarity 70.0%; Pred. No. 8;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 525 CLPGFSGDRC 534
I: |:|||||

RESULT 8
ID Q9GZ22 PRELIMINARY; PRT; 878 AA.
AC Q9GZ22;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Intestinal mucin (Fragment).
GN MUC3B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RA Kyo K., Parkes M., Ishigami H., Muto T., Nagawa H., Nakamura Y.;
RT "Human intestinal mucin gene 'MUC3' consists of two genes, MUC3A and
RT MUC3B: Associations of distinct variants of MUC3B with ulcerative
RT colitis and Crohn's disease.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB038783; BAB12117.1; -.
DR EMBL; AB038781; BAB12115.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER 1
SQ SEQUENCE 878 AA; 93685 MW; 1414E7E00AF5B4C CRC64;

Query Match      81.5%; Score 44; DB 4; Length 878;
Best Local Similarity 70.0%; Pred. No. 8;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 526 CLPGFSGDRC 535
I: |:|||||

RESULT 9
ID Q9H3Q7 PRELIMINARY; PRT; 878 AA.
AC Q9H3Q7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Intestinal mucin (Fragment).
GN MUC3A.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21183349; PubMed=11289722;
RA Kyo K., Muto T., Nagawa H., Lathrop GM., Nakamura Y.;
RT "Associations of distinct variants of the intestinal mucin gene MUC3A
RT with ulcerative colitis and Crohn's disease.";
RL J. Hum. Genet. 46:5-20(2001).
DR EMBL; AB038782; BAB12116.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF-like; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER 1
SQ SEQUENCE 878 AA; 93824 MW; B36E9B97F1EC5795 CRC64;

Query Match      81.5%; Score 44; DB 4; Length 878;
Best Local Similarity 70.0%; Pred. No. 8;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 526 CLPGFSGDRC 535
I: |:|||||

RESULT 10
ID Q9H195 PRELIMINARY; PRT; 901 AA.
AC Q9H195;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE MUC3B mucin (Fragment).
GN MUC3B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pratt W.S., Crawley S., Hicks J., Ho J., Nash M., Kim Y.S., Gum J.R.,
RA Swallow D.M.;
RT "Multiple transcripts of MUC3: Evidence for two genes, MUC3A and
RT MUC3B.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ291390; CAC19572.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER 1
SQ SEQUENCE 901 AA; 96192 MW; F98FC817494ECD99 CRC64;

Query Match      81.5%; Score 44; DB 4; Length 901;
Best Local Similarity 70.0%; Pred. No. 8.2;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 549 CLPGFSGDRC 558
I: |:|||||

```

```

RESULT 11
O14651
ID O14651 PRELIMINARY; PRT; 957 AA.
AC O14651: O14650;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Mucin (intestinal mucin) (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=97476275; PubMed=9334251;
RA Gum J.R. Jr., Ho J.J.L., Pratt W.S., Hicks J.W., Hill A.S.,
RA Vinall L.E., Robertson A.M., Swallow D.M., Kim Y.S.;
RT "MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl
RT terminus, and a novel upstream repetitive region.";
RL J. Biol. Chem. 272:26678-26686(1997).
DR EMBL: AF007194; AAC02272.1; -
DR EMBL: AF007196; AAB84383.1; -
DR EMBL: AF007198; AAB84382.1; -
DR InterPro: IPR000561; EGF-like.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 957 AA; 9554 MW; E67722560D600A1F CRC64;

Query Match 81.5%; Score 44; DB 4; Length 957;
Best Local Similarity 70.0%; Pred. No. 8.7;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 865 CLPGFSGDRC 874
I: I:|||||

RESULT 12
Q9UKW9
ID Q9UKW9 PRELIMINARY; PRT; 1217 AA.
AC Q9UKW9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Intestinal mucin 3 (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX TISSUE=INTESTINE;
RX MEDLINE=99443732; PubMed=10512748;
RA Crawley S.C., Gum J.R. Jr., Hicks J.W., Pratt W.S., Aubert J.P.,
RA Swallow D.M., Kim Y.S.;
RT "Genomic organization and structure of the 3' region of human MUC3:
RT alternative splicing predicts membrane-bound and soluble forms of the
RT mucin.";
RL Biochem. Biophys. Res. Commun. 263:728-736(1999).
DR EMBL: AF113616; AAF13032.1; -
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.

```

```

DR PROSITE; PS50024; SEA; 1.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 1217 AA; 129171 MW; 0783A4003C969664 CRC64;

Query Match 81.5%; Score 44; DB 4; Length 1217;
Best Local Similarity 70.0%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 865 CLPGFSGDRC 874
I: I:|||||

RESULT 13
Q28867
ID Q28867 PRELIMINARY; PRT; 89 AA.
AC Q28867;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor (Fragment).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=95000251; PubMed=7916972;
RA Stewart F., Power C.A., Lennard S.N., Allen W.R., Amet L.,
RA Edwards R.M.;
RT "Identification of the horse epidermal growth factor (EGF) coding
RT sequence and its use in monitoring EGF gene expression in the
RT endometrium of the pregnant mare.";
RL J. Mol. Endocrinol. 12:341-350(1994).
DR EMBL: S73527; AAB32226.1; -
DR HSSP; P01132; IEGF.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001336; EGF_1.
DR Pfam; PF000008; EGF; 2.
DR PRINTS; PR000009; EGFTGF.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00001; EGF_like; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 2.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 89 AA; 9879 MW; A7856F5E870B4A4B CRC64;

Query Match 79.6%; Score 43; DB 6; Length 89;
Best Local Similarity 60.0%; Pred. No. 1.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 74 CVGVGVGERC 83
I: I: I:|

RESULT 14
Q20043
ID Q20043 PRELIMINARY; PRT; 264 AA.
AC Q20043;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F35D2.3 protein.
GN F35D2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;

```

[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,  
RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Connell M.;  
RA "The sequence of C. elegans cosmid F35D2.";  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RN SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Waterston R.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U28741; AAA68325.1; -;  
DR HSP; P00740; IEDM.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR SMART; SM00180; EGF\_Lam; 1.  
DR SMART; SM00001; EGF\_Like; 1.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_5.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_1.  
KW EGF-like domain; Glycoprotein.  
SQ SEQUENCE 264 AA; 28434 MW; C91B1CC7B4B463F4 CRC64;

Query Match 79.6%; Score 43; DB 5; Length 264;

Best Local Similarity 70.0%; Pred. No. 3.5;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVIGXSGDRC 10

Db 76 CPVGLSGDRC 85

#### RESULT 15

Q90XX0  
ID Q90XX0 PRELIMINARY; PRT; 165 AA.  
AC Q90XX0;  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE New antigen receptor (Fragment).  
OS Orectolobus maculatus (spotted wobbegong).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Galeomorphi; Galeoidea; Orectolobiformes;  
OC Orectolobidae; Orectolobus.  
OX NCBI\_TaxID=168098;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=21451048; PubMed=11566324;  
RA Nuttall S.D., Krishnan U.V., Hattarki M., De Gori R., Irving R.A.,  
RA Hudson P.J.;  
RT "Isolation of the new antigen receptor from wobbegong sharks, and use  
as a scaffold for the display of protein loop libraries.";  
RL Mol. Immunol. 38:313-326(2001).  
DR EMBL; AF336090; AAK97358.1;  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR000504; RNA\_rec\_mot.

DR Pfam; PF00047; Ig; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
KW Receptor.  
FT NON\_TER. 165 165  
SQ SEQUENCE 165 AA; 18259 MW; D4A81143C6FDE1CD CRC64;  
Query Match 77.8%; Score 42; DB 13; Length 165;  
Best Local Similarity 77.8%; Pred. No. 3.4;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 VIGXSGDRC 10  
Db 105 MIGYSGDRC 113  
Search completed: July 2, 2003, 07:28:37  
Job time : 42.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 43.3333 Seconds  
(without alignments)  
30.750 Million cell updates/sec

Title: US-09-673-785D-9  
Perfect score: 56  
Sequence: 1 CVIGYSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

```

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred.. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	56	100.0	10	21	AAV52143
2	56	100.0	47	23	AAV52143
3	56	100.0	53	5	AAV52143
4	56	100.0	53	10	AAV52143
5	56	100.0	53	11	AAV52143
6	56	100.0	53	16	AAV52143
7	56	100.0	53	18	AAV52143
8	56	100.0	53	18	AAV52143
9	56	100.0	53	19	AAV52143
10	56	100.0	53	19	AAV52143

11	56	100.0	53	19	AAV50140
12	56	100.0	53	22	AAV50140
13	56	100.0	53	22	AAV50140
14	56	100.0	54	23	AAV50140
15	56	100.0	54	23	AAV50140
16	56	100.0	56	7	AAV50140
17	56	100.0	117	19	AAV50140
18	56	100.0	376	22	AAV50140
19	56	100.0	376	22	AAV50140
20	56	100.0	444	22	AAV50140
21	56	100.0	576	22	AAV50140
22	56	100.0	576	22	AAV50140
23	56	100.0	1073	22	AAV50140
24	56	100.0	1080	6	AAV50140
25	51	91.1	2146	22	AAV50140
26	49	87.5	655	23	AAV50140
27	49	87.5	741	21	AAV50140
28	49	87.5	2409	12	AAV50140
29	47	83.9	14	18	AAV50140
30	46	82.1	46	23	AAV50140
31	46	82.1	48	15	AAV50140
32	46	82.1	48	20	AAV50140
33	46	82.1	71	9	AAV50140
34	46	82.1	73	9	AAV50140
35	46	82.1	298	23	AAV50140
36	46	82.1	754	18	AAV50140
37	45	80.4	17	13	AAV50140
38	45	80.4	17	18	AAV50140
39	45	80.4	22	18	AAV50140
40	45	80.4	22	18	AAV50140
41	45	80.4	35	22	AAV50140
42	45	80.4	47	23	AAV50140
43	45	80.4	48	21	AAV50140
44	45	80.4	48	21	AAV50140
45	45	80.4	49	10	AAV50140

#### ALIGNMENTS

RESULT 1  
AAV52143  
ID AAV52143 standard; peptide; 10 AA.

XX AAV52143;

XX 28-JAN-2000 (first entry)

XX Mouse EGF derived peptide for targeting laminin receptor.

XX Epidermal growth factor; EGF; laminin receptor; angiogenesis; medicament; wound healing; retinopathy of immaturity; metastatic cancer; candida infection; leishmania; trichomonas vaginalis.

XX Mus sp.

XX Key Location/Qualifiers

XX Modified-site 1 /note= "Acetyl-Cys (S-ACM)"

XX Modified-site 10 /note= "Cys (S-ACM)-NH2"

XX WO9954356-A1.

XX 28-OCT-1999.

XX 21-APR-1999; 99WO-GB01211.

XX 22-APR-1998; 98GB-0008407.

XX (UYBE-) UNIV QUEBENS BELFAST.

XX Nelson J, Walker B, McFerran N, Harriott P;

Epidermal growth f  
Human EGF. Homo s  
Epidermal growth f  
Mouse epidermal gr  
Modified mouse epi  
Epidermal growth f  
Recombinant human  
Angiotensin conve  
Angiotensin conve  
Angiotensin conve  
Angiotensin conve  
Angiotensin conve  
Mouse epidermal gr  
Drosophila melanog  
Human versican iso  
Human cancer assoc  
Versican. Homo sa  
Chimeric epidermal  
Mouse NRG-2 EGF-11  
N terminal of rat  
Rat epithelial gro  
Rat epidermal grow  
Rat epidermal grow  
Human Neuregulin-2  
Rat cerebellum der  
EGF fragment. Syn  
Target peptide fro  
Target peptide fro  
Epidermal Growth F  
Human EGF motif se  
Nicked EGF1-48. H  
EGF-like domain of  
Epidermal growth f

XX WPI; 2000-013229/01.  
 XX New peptide derived from murine epidermal growth factor (mEGF)  
 XX Claim 4; Page 28; 35pp; English.  
 XX This is a peptide derived from mouse epidermal growth factor (EGF)  
 XX residues 33-42. This peptide is used in the invention to prepare a  
 XX composition to target laminin receptors. EGF derived peptides inhibit  
 XX blood vessel formation through their antagonism of the high affinity 67kD  
 XX laminin receptor found on endothelial cells. The peptide is modified  
 XX from the natural sequence to prevent protease attack. The peptide is used  
 XX in the preparation of a medicament for binding to laminin receptors as an  
 XX (ant)agonist. The medicament is also useful for healing endothelial cell  
 XX wounds and treating angiogenic diseases, especially retinopathy of  
 XX immaturity. Other diseases treated include metastatic cancer.  
 XX Candida spp. infection, and parasitic infestations like leishmania and  
 XX trichomonas vaginalis. The peptide are anti-angiogenic in human models.  
 XX The peptides also inhibit both laminin- and EGF-stimulated angiogenesis,  
 XX and prevent tumour cell attachment to basement membranes.  
 XX SQ Sequence 10 AA;  
 Query Match 100.0%; Score 56; DB 21; Length 10;  
 Best Local Similarity 90.0%; Pred. NO. 0.064;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGYSGDXC 10  
 Db 1 CVIGYSGDRC 10  
 |||||:|  
 |||||:|  
 RESULT 2  
 AAG6047  
 ID AAG66047 standard; peptide; 47 AA.  
 AC AAG66047;  
 XX 27-FEB-2002 (first entry)  
 DE Mouse EGF motif sequence.  
 XX ErbB-4; neuregulin-4; NRG-4; pro-NRG-4; neuroprotective; vulnery;  
 KW cerebroprotective; vasotropic; antiparkinsonian; anticonvulsant;  
 KW cytostatic; nootropic; EGF; NRG-2.  
 XX Mus musculus.  
 XX WO200181540-A2.  
 XX 01-NOV-2001.  
 XX 20-APR-2001; 2001WO-IL00371.  
 XX 21-APR-2000; 2000US-0553769.  
 XX (YEDA ) YEDA RES & DEV CO LTD.  
 XX Harari D, Yarden Y;  
 XX WPI; 2002-041398/05.  
 XX Novel ErbB-4 ligand, referred as neuregulin (NRG)-4 and polynucleotide  
 XX sequences encoding NRG-4, useful for upregulating or downregulating  
 XX ErbB-4 receptor activity to treat Alzheimer's disease, stroke, gastric  
 XX cancer  
 XX Disclosure; Fig 1c; 153pp; English.  
 XX The invention relates to a novel ErbB-4 ligand, neuregulin-4 (NRG-4).  
 XX NRG-4 binds to mammalian ErbB-4 receptor and can be expressed by standard  
 XX recombinant methodology. Pharmaceutical compositions comprising NRG-4 are

CC useful for regulating an endogenous protein affecting ErbB-4 receptor  
 CC activity in vivo. They are also useful for treating or preventing a  
 CC disease condition or syndrome associated with dysregulation of an  
 CC endogenous protein affecting ErbB-4 receptor activity, e.g., amyotrophic  
 CC lateral sclerosis (Lou Gehrig's disease), Bell's palsy, spinal muscular  
 CC atrophy, brain trauma, stroke, ischemia, Alzheimer's disease, Parkinson's  
 CC disease, epilepsy, multiple sclerosis, Huntington's chorea, Down's  
 CC syndrome, nerve deafness, neuropathy, muscular dystrophy, extramammary  
 CC Paget's disease, gastric, pancreatic, prostate, breast and ovarian  
 CC cancer, cervical carcinoma, endometrial adenocarcinoma, pancreatic D  
 CC cells-somatostatinoma and Zollinger-Ellison syndrome. The agent comprised  
 CC in the pharmaceutical composition includes a polypeptide (e.g., a soluble  
 CC ligand binding domain of ErbB-4 i.e., IgB4; or a monoclonal, polyclonal,  
 CC humanized, single chain antibody or an immunoreactive derivative of an  
 CC antibody) capable of binding the endogenous protein affecting ErbB-4  
 CC receptor activity. Traceable synthetic/recombinant NRG-4-tagged molecules  
 CC can serve as a diagnostic tool in which cells binding NRG-4 can be  
 CC measured. Sequences AAG66044-53 represent the EGF-like motifs of various  
 CC growth factors.  
 XX SQ Sequence 47 AA;  
 Query Match 100.0%; Score 56; DB 23; Length 47;  
 Best Local Similarity 90.0%; Pred. NO. 0.25;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGYSGDXC 10  
 Db 33 CVIGYSGDRC 42  
 |||||:|  
 |||||:|  
 RESULT 3  
 AAP40315  
 ID AAP40315 standard; protein; 53 AA.  
 AC AAP40315;  
 XX 22-JUL-1992 (first entry)  
 DE Epidermal growth factor tripentacontapeptide.  
 XX EGF.  
 XX Synthetic.  
 XX Key Location/Qualifiers  
 XX Disulfide-bond 6..20  
 XX Disulfide-bond 14..31  
 XX Disulfide-bond 33..42  
 XX JPS9027858-A.  
 XX 14-FEB-1984.  
 XX 05-AUG-1982; 82JP-0137128.  
 XX 05-AUG-1982; 82JP-0137128.  
 XX (NNSH ) NIPPON SHINYAKU KK.  
 XX WPI; 1984-072465/12.  
 XX Synthesis of epidermal growth factor polypeptide - by condensation  
 XX of protected smaller peptide sequences, de:protection then oxidn. to  
 XX cyclise.  
 XX Claim1; Page 1; 8pp; Japanese.  
 XX The amino acid sequence is that of an epidermal growth factor  
 XX tripentacontapeptide which is synthesised by condensation of  
 XX protected smaller peptide sequences. This method produces the  
 XX peptide smoothly, with high purity and yield.



SQ Sequence 53 AA;

Query Match 100.0%; Score 56; DB 5; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.28;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
DB 33 CVIGYSGDRC 42  
|||||||:|

#### RESULT 4

AAP91658  
ID AAP91658 standard; protein; 53 AA.

XX AC AAP91658;

DT 29-JUN-1990 (first entry)

DE Amino acid sequence for naturally occurring epidermal growth factor (EGF).

KW Epidermal growth factor; angiogenesis; synthetic peptide.

XX PN W08901489-A.

XX PD 23-FEB-1989.

PF 10-AUG-1988; 88WO-AU00300.

XX PR 10-AUG-1987; 87AU-0003629.

XX PA (CSIR ) COMMONWEALTH SCIENT ORG.

XX PI McAuslan BR;

XX DR WPI; 1989-068852/09.

PT Synthetic peptide active in stimulating angiogenesis -  
PT has sequences corresponding to amino acid sequences occurring in  
PT epidermal growth factor.

XX Fig 1; 1/1; 11pp; English.

XX The inventors claim synthetic peptides which correspond to sequences  
CC occurring in EGF, but excluding EGF. The synthetic peptides  
CC correspond to active sequences from EGF fragments: 3-14, 12-14, 12-15,  
CC 3-10, 29-37 and 33-37 (Claims 2-7 and AAP91659-P91664 resp.). The  
CC peptides are angiogenic. Their relative shortness means that they pose  
CC fewer synthesis problems than the entire EGF molecule.

XX SQ Sequence 53 AA;

Query Match 100.0%; Score 56; DB 10; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.28;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
DB 33 CVIGYSGDRC 42  
|||||||:|

#### RESULT 5

AAR08007  
ID AAR08007 standard; protein; 53 AA.

XX AC AAR08007;

DT 25-FEB-1991 (first entry)

DE Modified murine epidermal growth factor.

XX KW Modified murine epidermal growth factor; stability; storage;

KW epithelial wounds; gastric acid secretion.

XX Key Location/Qualifiers  
FT Misc-difference 11...11  
FT /label= E, N, Q, A, K

XX PN W09013570-A.

XX PD 15-NOV-1990.

XX PF 09-MAY-1990; 90WO-US02600.

XX PR 12-MAY-1989; 89US-0351773.

XX PA (CHIR-) CHIRON CORP.

XX PI Nascimento CG, Medina-Selby A;

XX DR WPI; 1990-361427/48.

XX Human epidermal growth factor - is substituted at position 11 for  
PT greater stability and improved storage life.

XX PS Claim 9; Page 25; 32pp; English.

XX The human EGF is used to treat oversecretion of gastric acid or an  
CC epithelial wound. EGF is modified to increase its chemical  
CC stability. Its storage life is improved without diminishing its  
CC biological activity. The proteins may be prepared by traditional  
CC chemical or recombinant means.  
CC See also AAR08004.

XX SQ Sequence 53 AA;

Query Match 100.0%; Score 56; DB 11; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.28;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
DB 33 CVIGYSGDRC 42  
|||||||:|

#### RESULT 6

AAR67275  
ID AAR67275 standard; peptide; 53 AA.

XX AC AAR67275;

XX DT 01-AUG-1995 (first entry)

XX DE Synthetic human/mouse EGF.

XX Cell growth factor; viscoelastic solution; fibroblastic growth factor;  
KW EGF; epidermal growth factor; EGF; buffered solution; lubrication;  
KW carboxypropyl methylcellulose; cellulose gum; dextran; dextran sulphate;  
KW chondroitin sulphate; sodium hyaluronate; osmolality; mitogenic;  
KW wound healing; cell protection; cell coating; surgery; tissue space;  
KW hydroxypropyl methylcellulose; manipulation.

XX OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 37  
FT /note= "Given in the specification as Try"

XX PN US5366964-A.

XX PD 22-NOV-1994.

XX PF 15-DEC-1988; 88US-0284533.

XX PR 15-DEC-1988; 88US-0284533.



PT Chimeric epidermal growth factor proteins - and DNA molecules for  
 XX their recombinant production  
 PS Claim 6; Page 26; 40pp; English.  
 XX This is a chimeric epidermal growth factor (EGF) protein ABC. This  
 CC chimeric EGF protein contains sequences derived from human and mouse  
 CC EGF sequences. The chimeric EGF proteins can be recombinantly produced  
 CC by culturing a microorganism of the genus *Bacillus* transformed by an  
 CC expression vector containing the encoding DNA molecules. The recombinant  
 CC chimeric EGF proteins exhibit a wide variety of physiological activities  
 CC similar to those of naturally occurring EGF.  
 XX  
 SQ Sequence 53 AA;  
 Query Match 100.0%; Score 56; DB 18; Length 53;  
 Best Local Similarity 90.0%; Pred. No. 0.28;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGYSGDXC 10  
 Db 33 CVIGYSGDRC 42  
 |||||:|  
 RESULT 9  
 AAW50134  
 ID AAW50134 standard; protein; 53 AA.  
 XX  
 AC AAW50134;  
 XX  
 DT 08-JUL-1998 (first entry)  
 XX  
 DE Mouse epidermal growth factor.  
 XX  
 KW Mouse; epidermal growth factor; EGF; isolation; recombinant;  
 KW ion exchange chromatography; *Bacillus brevis*.  
 XX  
 OS Mus sp.  
 XX  
 PN AU9728698-A.  
 XX  
 PD 05-FEB-1998.  
 XX  
 PF 17-JUL-1997; 97AU-0028698.  
 XX  
 PR 01-AUG-1996; 96JP-0218109.  
 XX  
 PA (HGET ) HIGETA SHOYU KK.  
 XX  
 PI Miyauchi A, Nemoto A, Warren B;  
 XX  
 DR WPI; 1998-131057/13.  
 XX  
 PT Isolation of recombinant epidermal growth factor from whole broth -  
 PT by fluidised-bed ion-exchange chromatography  
 XX  
 PS Claim 6; Page 26; 38pp; English.  
 XX  
 CC The present sequence represents mouse epidermal growth factor (EGF)  
 CC from the present invention. The present invention describes a method for  
 CC the isolation of recombinant epidermal growth factor (EGF) from whole  
 CC broth. The method comprises: (a) passing a culture containing the  
 CC protein upwards through a column containing a fluidised bed of ion  
 CC exchanger, washing the ion exchanger to remove non-adsorbed material,  
 CC allowing the ion exchanger to settle, and eluting adsorbed protein by  
 CC passing an eluant downwards through the column; (b) filtering the eluate  
 CC through a membrane with a molecular weight cutoff of 10000, and (c)  
 CC concentrating the filtrate on a membrane with a molecular weight cutoff  
 CC of 5000. The method is for isolating recombinant human, mouse, pig or  
 CC rat EGF, or an EGF analogue with one of four defined amino acid sequences  
 CC (given in the specification), from *Bacillus brevis* cultures. The EGF  
 CC proteins can be recovered in high yield and high purity without the need  
 CC to pretreat the culture broth to remove the cells, e.g. by costly

CC centrifugation or membrane filtration.  
 XX  
 SQ Sequence 53 AA;  
 Query Match 100.0%; Score 56; DB 19; Length 53;  
 Best Local Similarity 90.0%; Pred. No. 0.28;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGYSGDXC 10  
 Db 33 CVIGYSGDRC 42  
 |||||:|  
 RESULT 10  
 AAW50139  
 ID AAW50139 standard; protein; 53 AA.  
 XX  
 AC AAW50139;  
 XX  
 DT 08-JUL-1998 (first entry)  
 XX  
 DE Epidermal growth factor analogue 3.  
 XX  
 KW Epidermal growth factor; EGF; isolation; recombinant;  
 KW ion exchange chromatography; *Bacillus brevis*.  
 XX  
 OS Unidentified.  
 XX  
 PN AU9728698-A.  
 XX  
 PD 05-FEB-1998.  
 XX  
 PF 17-JUL-1997; 97AU-0028698.  
 XX  
 PR 01-AUG-1996; 96JP-0218109.  
 XX  
 PA (HGET ) HIGETA SHOYU KK.  
 XX  
 PI Miyauchi A, Nemoto A, Warren B;  
 XX  
 DR WPI; 1998-131057/13.  
 XX  
 PT Isolation of recombinant epidermal growth factor from whole broth -  
 PT by fluidised-bed ion-exchange chromatography  
 XX  
 PS Claim 6; Page 28; 38pp; English.  
 XX  
 CC The present sequence represents a protein which has epidermal growth  
 CC factor (EGF) like activity, from the present invention. The present  
 CC invention describes a method for the isolation of recombinant epidermal  
 CC growth factor (EGF) from whole broth. The method comprises: (a) passing  
 CC a culture containing the protein upwards through a column containing a  
 CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-  
 CC adsorbed material, allowing the ion exchanger to settle, and eluting  
 CC adsorbed protein by passing an eluant downwards through the column; (b)  
 CC filtering the eluate through a membrane with a molecular weight cutoff  
 CC of 10000, and (c) concentrating the filtrate on a membrane with a  
 CC molecular weight cutoff of 5000. The method is for isolating recombinant  
 CC human, mouse, pig or rat EGF, or an EGF analogue with one of four  
 CC defined amino acid sequences (given in the specification), from *Bacillus*  
 CC *brevis* cultures. The EGF proteins can be recovered in high yield and  
 CC high purity without the need to pretreat the culture broth to remove the  
 CC cells, e.g. by costly centrifugation or membrane filtration.  
 XX  
 SQ Sequence 53 AA;  
 Query Match 100.0%; Score 56; DB 19; Length 53;  
 Best Local Similarity 90.0%; Pred. No. 0.28;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGYSGDXC 10  
 Db 33 CVIGYSGDRC 42  
 |||||:|

```

RESULT 11
AAW50140
ID AAW50140 standard; protein; 53 AA.
XX
AC AAW50140;
XX
DT 08-JUL-1998 (first entry)
XX
DE Epidermal growth factor analogue 4.
XX
KW Epidermal growth factor; EGF; isolation; recombinant;
ion exchange chromatography; Bacillus brevis.
XX
OS Unidentified.
XX
PN AU9728698-A.
XX
PD 05-FEB-1998.
XX
PF 17-JUL-1997; 97AU-0028698.
XX
PR 01-AUG-1996; 96JP-0218109.
XX
PA (HGET ) HIGETA SHOYU KK.
XX
PI Miyauchi A, Nemoto A, Warren B;
XX
DR WPI; 1998-131057/13.
XX
PT Isolation of recombinant epidermal growth factor from whole broth -
PT by fluidised-bed ion-exchange chromatography
XX
PS Claim 6; Page 29; 38pp; English.
XX
CC The present sequence represents a protein which has epidermal growth
CC factor (EGF) like activity, from the present invention. The present
CC invention describes a method for the isolation of recombinant epidermal
CC growth factor (EGF) from whole broth. The method comprises: (a) passing
CC a culture containing the protein upwards through a column containing a
CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-
CC adsorbed material, allowing the ion exchanger to settle, and eluting
CC adsorbed protein by passing an eluent downwards through the column; (b)
CC filtering the eluate through a membrane with a molecular weight cutoff
CC of 10000, and (c) concentrating the filtrate on a membrane with a
CC molecular weight cutoff of 5000. The method is for isolating recombinant
CC human, mouse, pig or rat EGF, or an EGF analogue with one of four
CC defined amino acid sequences (given in the specification), from Bacillus
CC brevis cultures. The EGF proteins can be recovered in high yield and
CC high purity without the need to pretreat the culture broth to remove the
CC cells, e.g. by costly centrifugation or membrane filtration.
XX
SQ Sequence 53 AA;
Query Match 100.0%; Score 56; DB 19; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.28;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
DB 33 CVIGYSGDRC 42
|||||||:|

RESULT 12
AAW50140
ID AAW50140 standard; protein; 53 AA.
XX
AC AAW50140;
XX
DT 27-FEB-2001 (first entry)
XX
DE Human EGF.
XX

XX Merozoite surface protein; protozoacide; vaccine; malaria; human; EGF;
KW epidermal growth factor.
XX
OS Homo sapiens.
XX
PN WO200063245-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000WO-GB01558.
XX
PR 20-APR-1999; 99GB-0009072.
PR 13-MAY-1999; 99US-0311817.
PR 25-MAY-1999; 99CA-2271451.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiplbull C;
XX
DR WPI; 2001-015762/02.
XX
PT Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria -
XX
PS Disclosure; Fig 1; 126pp; English.
XX
CC The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of
CC the present invention are useful for immunising a mammal against malaria,
CC and can be used to treat malaria. The present sequence is human
CC epidermal growth factor (EGF). This sequence was used in a sequence
CC homology comparison with the wild-type MSP-1 protein from P. falciparum
CC (see AAB37608), which was used to generate the variants of the present
CC invention.
XX
SQ Sequence 53 AA;
Query Match 100.0%; Score 56; DB 22; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.28;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
DB 33 CVIGYSGDRC 42
|||||||:|

RESULT 13
AAE15714
ID AAE15714 standard; Protein; 53 AA.
XX
AC AAE15714;
XX
DT 12-MAR-2002 (first entry)
XX
DE Epidermal growth factor (EGF) used in exemplification of the invention.
XX
KW Humanised form; monoclonal antibody alpha 340; gene therapy;
KW epidermal growth factor receptor; EGF; cancer; colorectal; lung; breast;
KW gastric; ovarian; immune response; cytostatic; cell growth; apoptosis;
KW inhibitor.
XX
OS Unidentified.
XX
PN WO200188138-A1.
XX
PD 22-NOV-2001.

```

XX PF 21-MAY-2001; 2001WO-GB02226.  
 XX PR 19-MAY-2000; 2000GB-0011981.  
 XX PR 24-AUG-2000; 2000GB-0020794.  
 XX PA (SCAN-) SCANCELL LTD.  
 XX PI Ellis JRM, Durrant LG;  
 XX PI WPI; 2002-062384/08.  
 XX DR  
 XX PT New humanized form of mouse monoclonal antibody 340 which binds to  
 XX PT epidermal growth factor receptor and inhibits binding of growth factor,  
 XX PT useful for treating colorectal, lung, breast, gastric and ovarian  
 XX PT cancer.  
 XX PS Example 5; Fig 12; 53pp; English.  
 XX CC The present invention relates to a humanised form of the antibody 340 (a  
 CC mouse monoclonal antibody which binds to epidermal growth factor (EGF)  
 CC receptor and inhibits binding of EGF), obtainable from the cell line  
 CC deposited with the ECACC under accession number 97021428. The humanised  
 CC form of the antibody 340 is useful in gene therapy, medicine and in the  
 CC manufacture of a medicament for treatment or prophylaxis of cancer. The  
 CC invention is useful for treating colorectal, lung, breast, gastric or  
 CC ovarian cancers or also for preventing the recurrence of cancer after  
 CC initial treatment or surgery. The invention is also useful for enhancing  
 CC a protective immune response against cancer by optimised immunisation  
 CC schedules. The humanised form of the antibody 340 has reduced  
 CC immunogenicity but shows similar binding to cells expressing EGF  
 CC receptor, as the original murine antibody and has increased ability to  
 CC inhibit the growth of EGF receptor expressing cells. The invention is  
 CC used as cell growth and apoptosis inhibitor. The present sequence  
 CC is epidermal growth factor (EGF) which is used in the exemplification of  
 CC the invention.  
 XX SQ Sequence 53 AA;  
 Query Match 100.0%; Score 56; DB 23; Length 53;  
 Best Local Similarity 90.0%; Pred. No. 0.28;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGYSGDXC 10  
 Db 33 CVIGYSGDRC 42  
 RESULT 14  
 AAU76704  
 ID AAU76704 standard; Protein; 54 AA.  
 AC AAU76704;  
 XX 21-MAY-2002 (first entry)  
 XX Mouse epidermal growth factor (EGF).  
 XX DE  
 XX KW Mouse; epidermal growth factor; receptor; EGF;  
 KW chimeric polynucleotide; directed evolution; chimeraogenesis;  
 XX recombination.  
 XX OS Mus sp.  
 XX Key Location/Qualifiers  
 FT Misc-difference 54  
 FT /label= Unknown  
 FT /note= "Encoded by TAA"  
 XX  
 XX PN WO200206469-A2.  
 XX PD 24-JAN-2002.  
 XX

PF 18-JUL-2001; 2001WO-US22640.  
 XX PR 18-JUL-2000; 2000US-218921P.  
 XX PR 18-JUL-2000; 2000US-219085P.  
 XX PR 19-OCT-2000; 2000US-0691873.  
 XX PR 19-OCT-2000; 2000US-0692732.  
 XX PA (ENCH-) ENCHIRA BIOTECHNOLOGY CORP.  
 XX PI Coco WM, Ensell LP, Arensdorf JJ;  
 XX PI WPI; 2002-179795/23.  
 XX DR N-PSDB; ABK10313.  
 XX PT Forming chimeric polynucleotide by contacting single-stranded (ss)  
 XX PT scaffold fragment, and donor fragment populations to form hybridised  
 XX PT complex having ss regions which are filled-in, and ligating adjacent  
 XX PT fragments.  
 XX PS Example 2; Fig 3; 65pp; English.  
 XX CC The invention describes a method of forming (M1) chimeric  
 CC polynucleotides (I). The method comprises contacting a population of  
 CC single-stranded scaffold fragments (SF) with population of donor  
 CC fragments (DF) to form a complex (II) comprising at least one SF  
 CC hybridised to two DFs. (II) is treated so that single stranded regions of  
 CC (II) are filled-in and adjacent fragments are ligated. (M1) is useful for  
 CC a directed evolution process which involves forming a library of (I)  
 CC that can be screened for a characteristic of interest. Subsequent rounds  
 CC of directed evolution can produce chimeric polynucleotides with an  
 CC improved characteristic of interest. The methods facilitate the  
 CC generation of chimeric polynucleotides and do not require hybridising  
 CC donor fragments to a target- or full-length template. Because the  
 CC chimeraogenesis process does not rely upon a contiguous, full-length  
 CC template, it is unnecessary to modify a template to facilitate its  
 CC removal. This is the amino acid sequence of the mouse epidermal growth  
 CC factor (EGF) used to demonstrate a method of in vitro recombination  
 CC described in the method of the invention.  
 XX SQ Sequence 54 AA;  
 Query Match 100.0%; Score 56; DB 23; Length 54;  
 Best Local Similarity 90.0%; Pred. No. 0.28;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGYSGDXC 10  
 Db 33 CVIGYSGDRC 42  
 RESULT 15  
 AAU76706  
 ID AAU76706 standard; Protein; 54 AA.  
 AC AAU76706;  
 XX 21-MAY-2002 (first entry)  
 XX DE Modified mouse epidermal growth factor (EGF).  
 XX KW Mouse; epidermal growth factor; receptor; EGF;  
 KW chimeric polynucleotide; directed evolution; chimeraogenesis;  
 XX recombination.  
 XX OS Mus sp.  
 XX Key Location/Qualifiers  
 FT Misc-difference 38  
 FT /note= "Encoded by ACC"  
 XX  
 XX PN WO200206469-A2.  
 XX

PD 24-JAN-2002.  
XX  
XX PF 18-JUL-2001; 2001WO-US22640.  
XX  
XX PR 18-JUL-2000; 2000US-218921P.  
PR 18-JUL-2000; 2000US-219085P.  
PR 19-OCT-2000; 2000US-0691873.  
PR 19-OCT-2000; 2000US-0692732.  
XX  
XX PA (ENCH-) ENCHIRA BIOTECHNOLOGY CORP.  
XX  
XX PI Coco WM, Ensell LP, Arensdorf JJ;  
XX  
XX DR WPI; 2002-179795/23.  
DR N-PSDB; ABK10315.  
XX  
XX PT Forming chimeric polynucleotide by contacting single-stranded (ss)  
PT scaffold fragment, and donor fragment populations to form hybridised  
PT complex having ss regions which are filled-in, and ligating adjacent  
PT fragments -  
XX  
XX PS Example 2; Fig 3; 65pp; English.  
XX  
XX CC The invention describes a method of forming (M1) chimeric  
CC polynucleotides (I). The method comprises contacting a population of  
CC single-stranded scaffold fragments (SF) with population of donor  
CC fragments (DF) to form a complex (II) comprising at least one SF  
CC hybridised to two DFs. (II) is treated so that single stranded regions of  
CC (II) are filled-in and adjacent fragments are ligated. (M1) is useful for  
CC a directed evolution process which involves forming a library of (I)  
CC that can be screened for a characteristic of interest. Subsequent rounds  
CC of directed evolution can produce chimeric polynucleotides with an  
CC improved characteristic of interest. The methods facilitate the  
CC generation of chimeric polynucleotides and do not require hybridising  
CC donor fragments to a target- or full-length template. Because the  
CC chimeragenesis process does not rely upon a contiguous, full-length  
CC template, it is unnecessary to modify a template to facilitate its  
CC removal. This is the amino acid sequence of a mouse epidermal growth  
CC factor (EGF) modified to make it as similar as possible to the human  
CC protein (RAU76705) and used to demonstrate a method of in vitro  
CC recombination described in the method of the invention.  
XX  
XX SQ Sequence 54 AA;  
SQ

Query Match 100.0%; Score 56; DB 23; Length 54;  
Best Local Similarity 90.0%; Pred. No. 0.28;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CVIGYSGDXC 10  
| | | | | | | | | |  
Db 33 CVIGYSGDRC 42

Search completed: July 2, 2003, 07:30:55  
Job time : 44.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 10.6667 Seconds  
(without alignments)  
27.584 Million cell updates/sec

Title: US-09-673-785D-9  
Perfect score: 56  
Sequence: 1 CVIGYSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2.6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2.6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2.6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2.6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2.6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	12	1 US-08-597-545-7	Sequence 7, Appli
2	56	100.0	12	1 US-08-457-135-7	Sequence 7, Appli
3	56	100.0	53	1 US-08-284-923-2	Sequence 2, Appli
4	56	100.0	53	4 US-08-619-032B-2	Sequence 2, Appli
5	56	100.0	53	6 5332669-2	Patent No. 5332669
6	49	87.5	2409	6 5180808-2	Patent No. 5180808
7	48	85.7	154	3 US-09-191-647-10	Sequence 10, Appl
8	48	85.7	154	4 US-09-540-245A-10	Sequence 10, Appl
9	48	85.7	154	4 US-09-540-153-10	Sequence 10, Appl
10	46	82.1	48	6 5332669-1	Patent No. 5332669
11	46	82.1	754	2 US-08-525-864A-2	Sequence 2, Appli
12	45	80.4	17	6 5183805-1	Patent No. 5183805
13	45	80.4	17	6 5183805-2	Patent No. 5183805
14	45	80.4	25	6 5256643-8	Patent No. 5256643
15	45	80.4	37	2 US-08-039-364-14	Sequence 14, Appl
16	45	80.4	37	4 US-09-158-710-14	Sequence 14, Appl
17	45	80.4	42	1 US-08-168-091A-44	Sequence 44, Appl
18	45	80.4	44	1 US-08-278-089A-24	Sequence 24, Appl
19	45	80.4	46	3 US-08-899-437-12	Sequence 12, Appl
20	45	80.4	46	4 US-09-126-121-12	Sequence 12, Appl
21	45	80.4	47	3 US-08-753-007A-17	Sequence 17, Appl
22	45	80.4	47	4 US-09-398-496-17	Sequence 17, Appl
23	45	80.4	48	4 US-09-020-880-15	Sequence 15, Appl
24	45	80.4	48	4 US-08-915-096A-13	Sequence 13, Appl
25	45	80.4	48	4 US-09-101-544-15	Sequence 15, Appl
26	45	80.4	48	6 5434135-3	Patent No. 5434135
27	45	80.4	51	6 5177197-50	Patent No. 5177197

28	45	80.4	53	1 US-07-869-176-1	Sequence 1, Appli
29	45	80.4	53	1 US-08-284-923-1	Sequence 1, Appli
30	45	80.4	53	1 US-08-360-841-2	Sequence 2, Appli
31	45	80.4	53	2 US-08-861-000-1	Sequence 1, Appli
32	45	80.4	53	4 US-08-619-032B-1	Sequence 1, Appli
33	45	80.4	53	6 5332669-3	Patent No. 5332669
34	45	80.4	53	6 5434135-2	Patent No. 5434135
35	45	80.4	55	6 5218093-1	Patent No. 5218093
36	45	80.4	88	2 US-07-885-089B-13	Sequence 13, Appl
37	45	80.4	91	1 US-07-847-743B-15	Sequence 15, Appl
38	45	80.4	91	1 US-08-456-201-15	Sequence 15, Appl
39	45	80.4	91	2 US-08-330-161-13	Sequence 13, Appl
40	45	80.4	91	2 US-08-456-241-15	Sequence 15, Appl
41	45	80.4	91	2 US-08-440-401-13	Sequence 13, Appl
42	45	80.4	91	2 US-08-419-878B-13	Sequence 13, Appl
43	45	80.4	91	4 US-09-173-480-13	Sequence 13, Appl
44	45	80.4	91	5 PCT-US92-04295A-15	Sequence 15, Appl
45	45	80.4	93	1 US-08-343-401A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-08-597-545-7  
; Sequence 7, Application US/08597545  
; Patent No. 5580738  
; GENERAL INFORMATION:  
; APPLICANT: LABORDA, Jorge  
; TITLE OF INVENTION: Delta-Like Gene Expressed In  
; TITLE OF INVENTION: Neuroendocrine Tumors  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/597,545  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/989,537  
; FILING DATE: 11-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 40399/166 NIHD  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-597-545-7

Query Match 100.0%; Score 56; DB 1; Length 12;  
Best Local Similarity 90.0%; Pred. No. 0.011;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10

|||||||:|

Db 3 CVIGYSGDRC 12

RESULT 2  
US-08-457-135-7  
; Sequence 7, Application US/08457135  
; Patent No. 5644031  
; GENERAL INFORMATION:  
; APPLICANT: LABORDA, Jorge  
; TITLE OF INVENTION: Delta-Like Gene Expressed In  
; TITLE OF INVENTION: Neuroendocrine Tumors  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/457,135  
; FILING DATE: 01-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/989,537  
; FILING DATE: 11-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 40399/304/NIHD  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-457-135-7

Query Match 100.0%; Score 56; DB 1; Length 12;  
Best Local Similarity 90.0%; Pred. No. 0.011;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
| | | | | | | | | |  
DB 3 CVIGYSGDRC 12

RESULT 3  
US-08-284-923-2  
; Sequence 2, Application US/08284923  
; Patent No. 5547935  
; GENERAL INFORMATION:  
; APPLICANT: Mullenbach, Guy T  
; APPLICANT: Blaney, Jeffrey M  
; APPLICANT: Rosenberg, Steven  
; TITLE OF INVENTION: Muteins of Epidermal Growth Factor  
; TITLE OF INVENTION: exhibiting enhanced binding at low ph  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/284,923  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/746,651  
; FILING DATE: 16-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 231,001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-601-2708  
; TELEFAX: 510-655-3542  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 53 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-284-923-2

Query Match 100.0%; Score 56; DB 1; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.045;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
| | | | | | | | | |  
DB 33 CVIGYSGDRC 42

RESULT 4  
US-08-619-032B-2  
; Sequence 2, Application US/08619032B  
; Patent No. 6191106  
; GENERAL INFORMATION:  
; APPLICANT: Mullenbach, Guy T.  
; APPLICANT: Blaney, Jeffrey M.  
; APPLICANT: Rosenberg, Steven  
; TITLE OF INVENTION: MUTEINS OF EPIDERMAL GROWTH FACTOR  
; TITLE OF INVENTION: EXHIBITING ENHANCED BINDING AT LOW PH  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: Intellectual Property R-440, P.O. Box 8097  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94662-8097  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/619,032B  
; FILING DATE: 20-MAR-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Guth, Joseph H.  
; REGISTRATION NUMBER: 31,261  
; REFERENCE/DOCKET NUMBER: 0231,004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 923-3888  
; TELEFAX: (510) 655-3542  
; TELEX: N/A  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 53 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single



; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-619-032B-2

Query Match 100.0%; Score 56; DB 4; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.045;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
| | | | | | | | | |  
Db 33 CVIGYSGDRC 42

## RESULT 5

; Patent No. 5332669  
; APPLICANT: DEUEL, THOMAS F.  
; TITLE OF INVENTION: PROSTATE-DERIVED MITOGEN  
; NUMBER OF SEQUENCES: 3  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/356,739  
; FILING DATE: 24-MAR-1989  
; SEQ ID NO. 2:  
LENGTH: 53  
5332669-2

Query Match 100.0%; Score 56; DB 6; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.045;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
| | | | | | | | | |  
Db 33 CVIGYSGDRC 42

## RESULT 6

; Patent No. 5180808  
; APPLICANT: RUOSLAHTI, ERKKI I.  
; TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID  
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN  
; ANTIBODIES, AND METHODS OF DETECTING THE SAME  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/441,179  
; FILING DATE: 27-NOV-1989  
; SEQ ID NO. 2:  
LENGTH: 2409  
5180808-2

Query Match 87.5%; Score 49; DB 6; Length 2409;  
Best Local Similarity 80.0%; Pred. No. 19;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
| | | | | | | | | |  
Db 2128 CVPYSGDQC 2137

## RESULT 7

; Patent No. 6046015  
; APPLICANT: Goodman, Corey  
; APPLICANT: Kid, Thomas  
; APPLICANT: Brose, Katja  
; APPLICANT: Tessier-Lavigne, Marc  
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions  
; FILE REFERENCE: B98-031-3  
; CURRENT APPLICATION NUMBER: US/09/191,647  
; CURRENT FILING DATE: 1998-11-13  
; EARLIER APPLICATION NUMBER: 60/065,544

; EARLIER FILING DATE: 1997-11-14  
; EARLIER APPLICATION NUMBER: 60/081,057  
; EARLIER FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
LENGTH: 154  
; TYPE: PRT  
; ORGANISM: mouse  
US-09-191-647-10

Query Match 85.7%; Score 48; DB 3; Length 154;  
Best Local Similarity 60.0%; Pred. No. 1.9;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
| | | | | | | | | |  
Db 11 CMLGYTGDNC 20

## RESULT 8

; US-09-540-245A-10  
; Sequence 10, Application US/09540245A  
; Patent No. 6270984  
; GENERAL INFORMATION:  
; APPLICANT: Goodman, Corey  
; APPLICANT: Kid, Thomas  
; APPLICANT: Brose, Katja  
; APPLICANT: Tessier-Lavigne, Marc  
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions  
; FILE REFERENCE: B98-031-3  
; CURRENT APPLICATION NUMBER: US/09/540,245A  
; CURRENT FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 60/065,544  
; PRIOR FILING DATE: 1997-11-14  
; PRIOR APPLICATION NUMBER: 60/081,057  
; PRIOR FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
LENGTH: 154  
; TYPE: PRT  
; ORGANISM: mouse  
US-09-540-245A-10

Query Match 85.7%; Score 48; DB 4; Length 154;  
Best Local Similarity 60.0%; Pred. No. 1.9;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
| | | | | | | | | |  
Db 11 CMLGYTGDNC 20

## RESULT 9

; US-09-540-153-10  
; Sequence 10, Application US/09540153  
; Patent No. 6270995  
; GENERAL INFORMATION:  
; APPLICANT: Goodman, Corey  
; APPLICANT: Kid, Thomas  
; APPLICANT: Brose, Katja  
; APPLICANT: Tessier-Lavigne, Marc  
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions  
; FILE REFERENCE: B98-031-3  
; CURRENT APPLICATION NUMBER: US/09/540,153  
; CURRENT FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/191,647  
; PRIOR FILING DATE: 1998-11-13  
; PRIOR APPLICATION NUMBER: 60/081,057  
; PRIOR FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 10
; LENGTH: 154
; TYPE: PRT
; ORGANISM: mouse
US-09-540-153-10

Query Match      85.7%; Score 48; DB 4; Length 154;
Best Local Similarity 60.0%; Pred. No. 1.9;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 CVIGYSGDXC 10
Db      11 CMLGYTGDC 20
      |::|::|::|::|

RESULT 10
5332669-1
; Patent No. 5332669
; APPLICANT: DEUEL, THOMAS F.
; TITLE OF INVENTION: PROSTATE-DERIVED MITOGEN
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/356,739
; FILING DATE: 24-MAR-1989
; SEQ ID NO:1:
; LENGTH: 48
5332669-1

Query Match      82.1%; Score 46; DB 6; Length 48;
Best Local Similarity 70.0%; Pred. No. 1.3;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVIGYSGDXC 10
Db      33 CVIGYIGERC 42
      |::|::|::|::|

RESULT 11
US-08-525-864A-2
; Sequence 2, Application US/08525864A
; Patent No. 5912326
; GENERAL INFORMATION:
; APPLICANT: Chang, Han
; TITLE OF INVENTION: Cerebellum-derived Growth Factors, and Uses
; TITLE OF INVENTION: Related thereto
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,864A
; FILING DATE: 8-SEP-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kara, Catherine J.
; REGISTRATION NUMBER: 41,106
; REFERENCE/DOCKET NUMBER: HUI-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 754 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-864A-2

Query Match      82.1%; Score 46; DB 2; Length 754;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVIGYSGDXC 10
Db      280 CPVGYTGDC 289
      |::|::|::|::|

RESULT 12
5183805-1
; Patent No. 5183805
; APPLICANT: LEE, JIN S.;BLICK, MARK
; TITLE OF INVENTION: BIOACTIVE EGF PEPTIDES FOR
; PROMOTION OF TISSUE REGENERATION AND CANCER THERAPY
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/567,407
; FILING DATE: 13-AUG-1990
; SEQ ID NO:1:
; LENGTH: 17
5183805-1

Query Match      80.4%; Score 45; DB 6; Length 17;
Best Local Similarity 60.0%; Pred. No. 0.66;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVIGYSGDXC 10
Db      2 CVVGYIGERC 11
      |::|::|::|::|

RESULT 13
5183805-2
; Patent No. 5183805
; APPLICANT: LEE, JIN S.;BLICK, MARK
; TITLE OF INVENTION: BIOACTIVE EGF PEPTIDES FOR
; PROMOTION OF TISSUE REGENERATION AND CANCER THERAPY
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/567,407
; FILING DATE: 13-AUG-1990
; SEQ ID NO:2:
; LENGTH: 17
5183805-2

Query Match      80.4%; Score 45; DB 6; Length 17;
Best Local Similarity 60.0%; Pred. No. 0.66;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVIGYSGDXC 10
Db      2 CVVGYIGERC 11
      |::|::|::|::|

RESULT 14
5256643-8
; Patent No. 5256643
; APPLICANT: Persico, Maria G.;Salomon, David S.
; TITLE OF INVENTION: HUMAN CRYPTO PROTEIN
; NUMBER OF SEQUENCES: 18
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/530,165
; FILING DATE: 29-MAY-1990
; SEQ ID NO:8:
; LENGTH: 25
5256643-8

Query Match      80.4%; Score 45; DB 6; Length 25;
```

Best Local Similarity 60.0%; Pred. No. 0.96;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
Db 13 CVVGYIGERC 22

RESULT 15  
US-08-039-364-14  
; Sequence 14, Application US/08039364  
; Patent No. 5811393  
; GENERAL INFORMATION:  
; APPLICANT: Klagsbrun, Michael  
; APPLICANT: Abraham, Judith A.  
; APPLICANT: Higashiyama, Shigeki  
; APPLICANT: Besner, Gail F.  
; TITLE OF INVENTION: HEPARIN BINDING MITOGEN WITH  
; TITLE OF INVENTION: HOMOLOGY TO EPIDERMAL GROWTH FACTOR  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55Sx  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: Wordperfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/039,364  
; FILING DATE: 15 JUN 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/598,082  
; FILING DATE: 16 OCT 1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Freeman, John W.  
; REGISTRATION NUMBER: 29,066  
; REFERENCE/DOCKET NUMBER: 05162/002002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 37 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-039-364-14

Query Match 80.4%; Score 45; DB 2; Length 37;  
Best Local Similarity 60.0%; Pred. No. 1.4;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
Db 28 CVVGYIGERC 37

Search completed: July 2, 2003, 07:25:43  
Job time : 11 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:28:49 ; Search time 103 Seconds  
(without alignments)  
11.164 Million cell updates/sec

Title: US-09-673-785D-9  
Perfect score: 56  
Sequence: 1 CVIGYSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues  
Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	56	100.0	53	9	US-10-138-158-19
2	56	100.0	53	9	US-10-150-648B-33
3	56	100.0	145	9	US-10-150-648B-30
4	49	87.5	741	10	US-09-925-301-930
5	46	82.1	112	9	US-10-211-994-20
6	46	82.1	298	10	US-09-864-675-4
7	45	80.4	46	9	US-10-136-573A-12
8	45	80.4	46	9	US-09-877-665-12
9	45	80.4	46	9	US-10-215-862-12
10	45	80.4	46	10	US-09-817-647-12
11	45	80.4	47	12	US-10-096-241-17
12	45	80.4	48	9	US-10-201-945-13
13	45	80.4	53	9	US-09-903-327A-9
14	45	80.4	53	9	US-10-150-648B-35
15	45	80.4	53	9	US-10-211-994-4
16	45	80.4	53	10	US-09-848-664-31
17	45	80.4	58	10	US-09-934-706-3
18	45	80.4	91	9	US-10-022-609-13
19	45	80.4	111	9	US-10-211-994-27

20	45	80.4	111	9	US-10-211-994-29
21	45	80.4	112	9	US-10-211-994-5
22	45	80.4	112	9	US-10-211-994-7
23	45	80.4	112	9	US-10-211-994-25
24	45	80.4	140	10	US-09-280-030-64
25	45	80.4	146	9	US-10-150-648B-32
26	45	80.4	400	10	US-09-934-706-5
27	45	80.4	493	9	US-09-903-327A-13
28	45	80.4	832	9	US-09-783-931-6
29	45	80.4	832	10	US-09-908-322-6
30	44	78.6	176	9	US-10-024-599-4
31	44	78.6	572	9	US-09-900-449A-7
32	44	78.6	601	9	US-09-900-449A-5
33	44	78.6	639	9	US-09-900-449A-4
34	43	76.8	384	9	US-10-106-698-6253
35	42	75.0	43	9	US-10-050-704-269
36	42	75.0	53	9	US-10-150-648B-34
37	42	75.0	145	9	US-10-150-648B-31
38	42	75.0	211	9	US-10-050-704-266
39	42	75.0	427	9	US-10-050-704-273
40	42	75.0	483	9	US-10-050-704-272
41	42	75.0	566	9	US-09-900-449A-6
42	42	75.0	641	9	US-09-900-449A-8
43	42	75.0	774	9	US-09-900-449A-2
44	42	75.0	878	9	US-10-157-031-171
45	42	75.0	883	9	US-09-759-130B-342

ALIGNMENTS

RESULT 1  
US-10-138-158-19  
; Sequence 19, Application US/10138158  
; Publication No. US20030036509A1  
; GENERAL INFORMATION:  
; APPLICANT: STEM CELL PHARMACEUTICALS, INC.  
; APPLICANT: TWARDZIK, Daniel R.  
; APPLICANT: PERNET, Andre  
; APPLICANT: FELKER, Thomas S.  
; APPLICANT: PASKELL, Stefan  
; APPLICANT: RENO, John M.  
; TITLE OF INVENTION: TGF-alpha POLYPEPTIDES, FUNCTIONAL FRAGMENTS AND METHODS OF  
; FILE REFERENCE: STEM110-6  
; CURRENT APPLICATION NUMBER: US/10/138.158  
; CURRENT FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: US 09/641,587  
; PRIOR FILING DATE: 2000-08-17  
; PRIOR APPLICATION NUMBER: US 09/559,248  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: US 09/459,813  
; PRIOR FILING DATE: 1999-12-13  
; PRIOR APPLICATION NUMBER: US 09/492,935  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 09/378,567  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19  
; LENGTH: 53  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-138-158-19

Query Match 100.0%; Score 56; DB 9; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.064;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
| | | | | | | | | |  
Db 33 CVIGYSGDRC 42

RESULT 2  
US-10-150-648B-33  
; Sequence 33, Application US/10150648B  
; Publication No. US20030059802A1  
; GENERAL INFORMATION:  
; APPLICANT: Bilodeau-Goesseels, Sylvie  
; APPLICANT: John, Sushil J.  
; APPLICANT: Selinger, Leonard B.  
; APPLICANT: Benkel, Bernhard F.  
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth  
; TITLE OF INVENTION: factor  
; FILE REFERENCE: 60-01  
; CURRENT APPLICATION NUMBER: US/10/150,648B  
; CURRENT FILING DATE: 2002-05-17  
; PRIOR APPLICATION NUMBER: 60/292,136  
; PRIOR FILING DATE: 2001-05-18  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 33  
; LENGTH: 53  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; PUBLICATION INFORMATION:  
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.  
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a  
; TITLE: 128,000-molecular weight protein precursor  
; JOURNAL: Nature  
; VOLUME: 303  
; PAGES: 722-725  
; DATE: 1983  
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380  
; DATABASE ENTRY DATE: 1993-04-27  
; RELEVANT RESIDUES: Relevant residues FROM 977 TO 1029  
US-10-150-648B-33

Query Match 100.0%; Score 56; DB 9; Length 53;  
Best Local Similarity 90.0%; Pred. No. 0.064; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
|||||||  
DB 33 CVIGYSGDRC 42

RESULT 3  
US-10-150-648B-30  
; Sequence 30, Application US/10150648B  
; Publication No. US20030059802A1  
; GENERAL INFORMATION:  
; APPLICANT: Bilodeau-Goesseels, Sylvie  
; APPLICANT: John, Sushil J.  
; APPLICANT: Selinger, Leonard B.  
; APPLICANT: Benkel, Bernhard F.  
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth  
; TITLE OF INVENTION: factor  
; FILE REFERENCE: 60-01  
; CURRENT APPLICATION NUMBER: US/10/150,648B  
; CURRENT FILING DATE: 2002-05-17  
; PRIOR APPLICATION NUMBER: 60/292,136  
; PRIOR FILING DATE: 2001-05-18  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 30  
; LENGTH: 145  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; PUBLICATION INFORMATION:  
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.  
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a  
; TITLE: 128,000-molecular weight protein precursor  
; JOURNAL: Nature  
; VOLUME: 303  
; PAGES: 722-725

DATE: 1983  
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380  
; DATABASE ENTRY DATE: 1993-04-27  
; RELEVANT RESIDUES: Relevant residues FROM 919 TO 1063  
US-10-150-648B-30

Query Match 100.0%; Score 56; DB 9; Length 145;  
Best Local Similarity 90.0%; Pred. No. 0.15; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
|||||||  
DB 91 CVIGYSGDRC 100

RESULT 4  
US-09-925-301-930  
; Sequence 930, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 930  
; LENGTH: 741  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (8)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (19)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (282)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-301-930

Query Match 87.5%; Score 49; DB 10; Length 741;  
Best Local Similarity 80.0%; Pred. No. 7.2; Mismatches 1; Indels 0; Gaps 0;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
|||||||  
DB 460 CVPYSGDQC 469

RESULT 5  
US-10-211-994-20  
; Sequence 20, Application US/10211994  
; Publication No. US20030082201A1  
; GENERAL INFORMATION:  
; APPLICANT: Rao, M.R.S.  
; APPLICANT: Sengupta, Paromita  
; APPLICANT: Prasad, Sudhanand  
; APPLICANT: Burman, Anand C.  
; APPLICANT: Mukherjee, Rama  
; APPLICANT: Thomas, Becky  
; TITLE OF INVENTION: MULTIVALENT SYNTHETIC VACCINE FOR CANCER  
; FILE REFERENCE: U014152-1  
; CURRENT APPLICATION NUMBER: US/10/211,994  
; CURRENT FILING DATE: 2002-08-02  
; PRIOR APPLICATION NUMBER: 60/309,975  
; PRIOR FILING DATE: 2001-08-03

NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 112  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Nucleotide and peptide sequence of multivalent vaccine  
US-10-211-994-20

Query Match 82.1%; Score 46; DB 9; Length 112;  
Best Local Similarity 70.0%; Pred. No. 4.1;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
Db 92 CVVGYIGERC 101

## RESULT 6

US-09-864-675-4  
; Sequence 4, Application US/09864675  
; Patent No. US20020081286A1  
; GENERAL INFORMATION:  
; APPLICANT: Marchionni, Mark  
; TITLE OF INVENTION: NRG-2 NUCLEIC ACID MOLECULES,  
; POLYPEPTIDES, AND DIAGNOSTIC AND THERAPEUTIC METHODS  
; FILE REFERENCE: 04595/049002  
; CURRENT APPLICATION NUMBER: US/09/864,675  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/206,495  
; PRIOR FILING DATE: 2000-05-23  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 298  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-864-675-4

Query Match 82.1%; Score 46; DB 10; Length 298;  
Best Local Similarity 60.0%; Pred. No. 9.5;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
Db 280 CPVGYTGDCR 289

## RESULT 7

US-10-136-573A-12  
; Sequence 12, Application US/10136573A  
; Patent No. US20020161200A1  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Mark, Melanie Rose  
; APPLICANT: Zhang, Dong Xiao  
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and  
; TITLE OF INVENTION: Uses Therefor  
; FILE REFERENCE: P1084R1C2  
; CURRENT APPLICATION NUMBER: US/10/136,573A  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: US 09/480,977  
; PRIOR FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: US 08/899,437  
; PRIOR FILING DATE: 1997-07-24  
; PRIOR APPLICATION NUMBER: US 60/052,019  
; PRIOR FILING DATE: 1997-07-09  
; NUMBER OF SEQ ID NOS: 23  
; SEQ ID NO 12  
; LENGTH: 46  
; TYPE: PRT  
; ORGANISM: Homo sapiens

## US-10-136-573A-12

Query Match 80.4%; Score 45; DB 9; Length 46;  
Best Local Similarity 60.0%; Pred. No. 2.7;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
Db 32 CVVGYIGERC 41

## RESULT 8

US-09-877-665-12  
; Sequence 12, Application US/09877665  
; Patent No. US20020164680A1  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao  
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related  
; Ligands and Uses Therefor  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/877,665  
FILING DATE: 08-Jun-2001  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/109,206  
FILING DATE: 30-Jun-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Conley, Delidre L.  
REGISTRATION NUMBER: 36,487  
REFERENCE/DOCKET NUMBER: P1084R1-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-2066  
TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear

FEATURE:  
NAME/KEY: hEGF.egf  
LOCATION: 1-46  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
SEQUENCE DESCRIPTION: SEQ ID NO: 12:

US-09-877-665-12

Query Match 80.4%; Score 45; DB 9; Length 46;  
Best Local Similarity 60.0%; Pred. No. 2.7;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10  
Db 32 CVVGYIGERC 41

## RESULT 9

US-10-215-862-12  
; Sequence 12, Application US/10215862  
; Publication No. US20030036166A1  
; GENERAL INFORMATION:

```
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie Rose
; APPLICANT: Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: PI084R1d2C1
; CURRENT APPLICATION NUMBER: US/10/215,862
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 09/126,663
; PRIOR FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: US 08/899,437
; PRIOR FILING DATE: 1997-07-24
; PRIOR APPLICATION NUMBER: US 60/052,019
; PRIOR FILING DATE: 1997-07-09
; NUMBER OF SEQ ID NOS: 23
; SEQ ID NO 12
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-215-862-12

Query Match      80.4%; Score 45; DB 9; Length 46;
Best Local Similarity 60.0%; Pred. No. 2.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVVIGSGDNC 10
        ||:|| |::|
Db      32 CVVGYIGERC 41

RESULT 10
US-09-817-647-12
; Sequence 12, Application US/09817647
; Patent No. US2002082229A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/817,647
; FILING DATE: 26-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/107,979
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: PI084R1-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: hEGF.egf
; LOCATION: 1-46
```

```
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-817-647-12

Query Match      80.4%; Score 45; DB 10; Length 46;
Best Local Similarity 60.0%; Pred. No. 2.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVVIGSGDNC 10
        ||:|| |::|
Db      32 CVVGYIGERC 41

RESULT 11
US-10-096-241-17
; Sequence 17, Application US/10096241
; Patent No. US20020127594A1
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/096,241
; FILING DATE: 12-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 47 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-096-241-17

Query Match      80.4%; Score 45; DB 12; Length 47;
Best Local Similarity 60.0%; Pred. No. 2.8;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVVIGSGDNC 10
        ||:|| |::|
Db      33 CVVGYIGERC 42

RESULT 12
US-10-201-945-13
; Sequence 13, Application US/10201945
```



Publication No. US20020188110A1  
GENERAL INFORMATION:  
APPLICANT: Weissner, Paul S.  
Fuldner, Rebecca A.  
Adams, Mark D.  
TITLE OF INVENTION: Transforming Growth Factor Alpha HI  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/201,945  
FILING DATE: 25-Jul-2002  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/09/471,377  
FILING DATE: 23-Dec-1999  
APPLICATION NUMBER: 08/915,096  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 08/208,008  
FILING DATE: 08-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF110D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8504  
TELEX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-10-201-945-13

Query Match 80.4%; Score 45; DB 9; Length 48;  
Best Local Similarity 60.0%; Pred. No. 2.8;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10  
DB 28 CVVGIIGERC 37

RESULT 13  
US-09-903-327A-9  
Sequence 9, Application US/09903327A  
Patent No. US2002016433A1  
GENERAL INFORMATION:  
APPLICANT: Nemerow, Glen R.  
Li, Erquang  
TITLE OF INVENTION: BIFUNCTIONAL MOLECULES AND VECTORS COMPLEXED THEREWITH FOR TARGET  
TITLE OF INVENTION: GENE  
TITLE OF INVENTION: DELIVERY  
FILE REFERENCE: 22908-1228  
CURRENT APPLICATION NUMBER: US/09/903,327A  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 09/613,017  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 9  
LENGTH: 53  
TYPE: PRT  
ORGANISM: Human  
FEATURE:  
NAME/KEY: PEPTIDE  
LOCATION: (0)...(0)  
OTHER INFORMATION: Epidermal Growth Factor (EGF, mature peptide)  
US-09-903-327A-9

Query Match 80.4%; Score 45; DB 9; Length 53;  
Best Local Similarity 60.0%; Pred. No. 3.1;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10  
DB 33 CVVGIIGERC 42

RESULT 14  
US-10-150-648B-35  
Sequence 35, Application US/10150648B  
Publication No. US20030059802A1  
GENERAL INFORMATION:  
APPLICANT: Bilodeau-Goesseels, Sylvie  
APPLICANT: John, Sushil J.  
APPLICANT: Selinger, Leonard B.  
APPLICANT: Benkel, Bernhard F.  
TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth  
factor  
FILE REFERENCE: 60-01  
CURRENT APPLICATION NUMBER: US/10/150,648B  
CURRENT FILING DATE: 2002-05-17  
PRIOR APPLICATION NUMBER: 60/292,136  
PRIOR FILING DATE: 2001-05-18  
NUMBER OF SEQ ID NOS: 37  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 35  
LENGTH: 53  
TYPE: PRT  
ORGANISM: Homo sapiens  
PUBLICATION INFORMATION:  
AUTHORS: Bell, G.I., Fong, N.M., Stempfen, M.M., Wormsted, M.A.,  
AUTHORS: Caput, D., Ku, L.L., Urdea, M.S., Rall, L.B. and Sanchez-  
AUTHORS: Pescador, R.  
TITLE: Human epidermal growth factor precursor: cDNA sequence,  
TITLE: expression in vitro and gene organization.  
JOURNAL: Nucleic Acids Research  
VOLUME: 14  
ISSUE: 21  
PAGES: 8427-8446  
DATE: 1986  
DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 X04571  
DATABASE ENTRY DATE: 1993-04-21  
RELEVANT RESIDUES: Relevant residues FROM 970 TO 1022  
US-10-150-648B-35

Query Match 80.4%; Score 45; DB 9; Length 53;  
Best Local Similarity 60.0%; Pred. No. 3.1;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10  
DB 33 CVVGIIGERC 42

RESULT 15  
US-10-211-994-4  
Sequence 4, Application US/10211994  
Publication No. US20030082201A1  
GENERAL INFORMATION:  
APPLICANT: Rao, M.R.S.  
APPLICANT: Sengupta, Paromita

```
; APPLICANT: Prasad, Sudhanand
; APPLICANT: Burman, Anand C.
; APPLICANT: Mukherjee, Rama
; APPLICANT: Thomas, Becky
; TITLE OF INVENTION: MULTIVALENT SYNTHETIC VACCINE FOR CANCER
; FILE REFERENCE: U014152-1
; CURRENT APPLICATION NUMBER: US/10/211,994
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/309,975
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Epidermal growth factor
US-10-211-994-4

Query Match      80.4%; Score 45; DB 9; Length 53;
Best Local Similarity 60.0%; Pred. NO. 3.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVIGYSGDXC 10
        ||:||| |::|
Db      33 CVVGIGERC 42

Search completed: July 2, 2003, 07:57:18
Job time : 103 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 12 Seconds  
(without alignments)  
80.112 Million cell updates/sec

Title: US-09-673-785D-9

Perfect score: 56

Sequence: 1 CVIGYSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	1217	1 EGSMG	epidermal growth f
2	51	91.1	2139	2 A35672	crumbs protein - f
3	51	91.1	4006	2 T09070	probable tenascin
4	49	87.5	102	2 B5885	Chondrotin sulfat
5	49	87.5	372	2 T23559	hypothetical prote
6	49	87.5	862	2 S43922	versican - pig-tai
7	49	87.5	1643	2 T14274	versican precursor
8	49	87.5	2409	1 A60979	versican precursor
9	49	87.5	3381	2 T42389	versican precursor
10	46	82.1	57	2 PC4415	ErbB kinase activa
11	46	82.1	1133	1 EGRT	epidermal growth f
12	46	82.1	1531	2 T42218	slit-1 protein hom
13	45	80.4	832	2 A31246	neurogenic protein
14	45	80.4	833	2 S19087	gene Delta protein
15	45	80.4	880	2 S00670	neurogenic repetit
16	45	80.4	883	2 S43126	brevican precursor
17	45	80.4	1207	1 EGHU	epidermal growth f
18	45	80.4	1722	2 E89753	protein Filic7.4 (i
19	45	80.4	2397	1 A55535	versican precursor
20	44	78.6	264	2 T16271	hypothetical prote
21	44	78.6	1064	2 A40136	fibropellin Ia - s
22	44	78.6	1203	2 A43175	Notch B protein -
23	44	78.6	2352	2 T30201	Notch homolog prot
24	44	78.6	2471	2 A49128	cell-fate determin
25	44	78.6	2476	2 T34022	zonadhesin - pig
26	44	78.6	3566	1 A40701	tenascin-X precurs
27	44	78.6	4135	2 T42629	tenascin-X - bovin
28	43	76.8	373	2 T23300	hypothetical prote
29	43	76.8	2321	2 S78549	notch3 protein - h

## ALIGNMENTS

## RESULT 1

## EGSMG

epidermal growth factor precursor - mouse

N;Alternate names: urogastrone precursor

C;Species: Mus musculus (house mouse)

C;Date: 30-Nov-1980 #sequence.Revision 11-Aug-1983 #text\_change 19-Jan-2001

C;Accession: A94272; A93304; A92118; A01387

R;Scott, J.; Urdea, M.; Quiroga, M.; Sanchez-Pescador, R.; Fong, N.; Selby, M.; Rutt

Science 221, 236-240, 1983

A;Title: Structure of a mouse submaxillary messenger RNA encoding epidermal growth f

A;Reference number: A94272; MUID:83223630; PMID:6602382

A;Accession: A94272

A;Molecule type: mRNA

A;Residues: 1-1217 <SCO>

A;Cross-references: GB:J00380; NID:g192993; PIDN:AAA37539.1; PID:g309210

R;Gray, A.; Dull, T.; Ullrich, A.

Nature 303, 722-725, 1983

A;Title: Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-mole

A;Reference number: A93304; MUID:83219309; PMID:6304537

A;Accession: A93304

A;Molecule type: mRNA

A;Residues: 1-789, 'Y', 791-1047, 'S', 1049-1168 <GRA>

A;Cross-references: GB:J00380

A;Note: the sequence shown by these authors differs from residues 1134-1168 due to a

ence of Scott et al.)

R;Savage Jr., C.R.; Inagami, T.; Cohen, S.

J. Biol. Chem. 247, 7612-7621, 1972

A;Title: The primary structure of epidermal growth factor.

A;Reference number: A92118; MUID:73048516; PMID:4636327

A;Accession: A92118

A;Molecule type: protein

A;Residues: 977-1029 <SAV>

A;Note: residues 1024-1029 are not required for full biological activity in vivo

R;Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A;Title: Epidermal growth factor. Location of disulfide bonds.

A;Reference number: A92144; MUID:74025498; PMID:4750422

A;Contents: annotation; disulfide bonds

C;Comment: Epidermal growth factor (EGF) stimulates the proliferation and differenti

gastrointestinal cell proliferation.

C;Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in f

C;Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound pr

C;Comment: The active growth factor from this submaxillary gland protein stimulates

C;Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-co

C;Keywords: duplication; growth factor; tandem repeat; transmembrane protein

F;1-28/Domain: signal sequence #status predicted <SIG>

F;29-1217/Product: epidermal growth factor proprotein, membrane-bound form #status p

F;29-1038/Domain: extracellular #status predicted <EXT>

F;50-485/Region: EGF precursor long repeat

F;53-92/Domain: LDL receptor YWTD-containing repeat homology <YW01>

F;93-134/Domain: LDL receptor YWTD-containing repeat homology <YW02>

F:135-176/Domain: LDL receptor WYTD-containing repeat homology <YW03>  
 F:177-217/Domain: LDL receptor WYTD-containing repeat homology <YW04>  
 F:218-262/Domain: LDL receptor WYTD-containing repeat homology <YW05>  
 F:263-307/Domain: LDL receptor WYTD-containing repeat homology <YW06>  
 F:324-360/Domain: EGF homology #status atypical <EG1>  
 F:366-401/Domain: EGF homology <EG2>  
 F:407-442/Domain: EGF homology <EG3>  
 F:445-482/Domain: EGF homology <EG4>  
 F:486-529/Region: EGF precursor long repeat  
 F:489-529/Domain: LDL receptor WYTD-containing repeat homology <YW07>  
 F:530-572/Domain: LDL receptor WYTD-containing repeat homology <YW08>  
 F:573-615/Domain: LDL receptor WYTD-containing repeat homology <YW09>  
 F:616-659/Domain: LDL receptor WYTD-containing repeat homology <YW10>  
 F:660-700/Domain: LDL receptor WYTD-containing repeat homology <YW11>  
 F:701-743/Domain: LDL receptor WYTD-containing repeat homology <YW12>  
 F:751-786/Domain: EGF homology <EG5>  
 F:842-875/Domain: EGF homology <EG6>  
 F:881-917/Domain: EGF homology <EG7>  
 F:923-958/Domain: EGF homology <EG8>  
 F:977-1029/Product: epidermal growth factor #status experimental <EGF>  
 F:982-1018/Domain: EGF homology <EG9>  
 F:1039-1063/Domain: transmembrane #status predicted <TMW>  
 F:1064-1217/Domain: intracellular #status predicted <INT>  
 F:347-360,366-377,373-386,388-401,407-418,414-427,429-442,445-457,453-467,469-482,751-786 predicted  
 F:982-996,990-1007,1009-1018/Disulfide bonds: #status experimental

Query Match 100.0%; Score 56; DB 1; Length 1217;  
 Best Local Similarity 90.0%; Pred. No. 0.12;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGYSGDXC 10  
 | : ||||| :  
 Db 1009 CVIGYSGDRC 1018

## RESULT 2

A35672  
 crumbs protein - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C>Date: 21-Sep-1990 #sequence\_revision 18-Nov-1992 #text\_change 11-Jan-2000  
 C:Accession: A35672  
 R:Tapas, U.; Theres, C.; Knust, E.  
 Cell 61, 787-799, 1990  
 A:Title: crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila  
 A:Reference number: A35672; MUID:90263104; PMID:2344615  
 A:Accession: A35672  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-2139 <TEP>  
 A:Cross-references: GB:M33753  
 A:Note: the authors translated the codon GGC for residue 1928 as Cys, and TAT for residue  
 C:Genetics:

A:Gene: FlyBase:crb  
 A:Cross-references: FlyBase:FBgn0000368  
 C:Superfamily: unassigned EGF-related proteins; EGF homology  
 C:Keywords: transmembrane protein  
 F:352-385/Domain: EGF homology <EGX1>  
 F:392-424/Domain: EGF homology <EGF1>  
 F:691-722/Domain: EGF homology <EGF>  
 F:767-799/Domain: EGF homology <EGF3>  
 F:1878-1914/Domain: EGF homology <EGX2>

Query Match 91.1%; Score 51; DB 2; Length 2139;  
 Best Local Similarity 70.0%; Pred. No. 1.5;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGYSGDXC 10  
 | : ||||| :  
 Db 571 CAVGYSGDRC 580

## RESULT 3

T09070  
 probable tenascin X - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jan-2000  
 C:Accession: T09070  
 R:Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.;  
 submitted to the EMBL Data Library, October 1997  
 A:Description: Sequence of the mouse major histocompatibility locus class III region  
 A:Reference number: 216543  
 A:Accession: T09070  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-4006 <ROW>  
 A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958  
 C:Genetics:  
 A:Gene: TNX  
 A:Map position: 17  
 A:Introns: 124/1; 735/1; 773/3; 826/1; 914/1; 1037/1; 1135/1; 1232/1; 1329/1; 1440/1;  
 019/1; 3115/1; 3208/1; 3302/1; 3405/1; 3517/1; 3558/1; 3606/1; 3646/1; 3694/1; 3730/3  
 C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin  
 C:Keywords: extracellular matrix  
 F:422-448/Domain: EGF homology <EGF>  
 F:826-906/Domain: fibronectin type III repeat homology <3FR>  
 F:3789-3997/Domain: fibrinogen beta/gamma homology <FBG>

Query Match 91.1%; Score 51; DB 2; Length 4006;  
 Best Local Similarity 70.0%; Pred. No. 2.7;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGYSGDXC 10  
 | : ||||| :  
 Db 532 CAVGYSGDDC 541

## RESULT 4

B55885  
 chondroitin sulfate proteoglycan MV3 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C>Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 16-Jul-1999  
 C:Accession: B55885  
 R:Zako, M.; Shinomura, T.; Ujita, M.; Ito, K.; Kimata, K.  
 J. Biol. Chem. 270, 3914-3918, 1995  
 A:Title: Expression of PG-M(V3), an alternatively spliced form of PG-M without a cnon  
 A:Reference number: A55885; MUID:95181355; PMID:7876137  
 A:Accession: B55885  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-102 <ZAK>  
 A:Cross-references: GB:S75879; GB:D32039  
 C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology;  
 F:24-55/Domain: EGF homology <EG1>  
 F:62-93/Domain: EGF homology <EG2>

Query Match 87.5%; Score 49; DB 2; Length 102;  
 Best Local Similarity 80.0%; Pred. No. 0.18;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGYSGDXC 10  
 | : ||||| :  
 Db 46 CVPGYSGDQC 55

## RESULT 5

T29359  
 hypothetical protein R05G6.9 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T29359  
 R:Murray, J.; Le, T.T.  
 submitted to the EMBL Data Library, May 1996  
 A:Description: The sequence of C. elegans cosmid R05G6.  
 A:Reference number: 220612  
 A:Accession: T29359

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-372 <EMBL>

A:Cross-references: EMBL:U58746; PIDN:AAB00626.1; GSPDB:GN00022; CESP:R05G6.9

A:Experimental source: strain Bristol N2; clone R05G6

C:Genetics:

A:Gene: CESP:R05G6.9

A:Map position: 4

A:Introns: 80/1; 161/1; 245/1; 286/1

Query Match 87.5%; Score 49; DB 2; Length 372;

Best Local Similarity 80.0%; Pred. No. 0.61;

Matches 7; Conservative 2; Mismatches 0; Gaps 0;

QY 1 CVIGYSGDXC 10

|| |||||:|

Db 233 CVLYSGDKC 242

#### RESULT 6

S43922

N:Alternate names: chondroitin sulfate proteoglycan

C:Species: Macaca nemestrina (pig-tailed macaque)

C>Date: 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 16-Jul-1999

C:Accession: S43922

R:Yao, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.

Matrix Biol. 14, 213-225, 1994

A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by

A:Reference number: S43921; MUID:95005762; PMID:7921538

A:Accession: S43922

A:Molecule type: mRNA

A:Residues: 1-233;234-525;526-862 <YAO>

A:Cross-references: EMBL:572413

A>Note: 507-Ser was also found

A>Note: the authors translated the codon GCC for residue 50 as Val, AAG for residue 422

669 as Asn

C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; E

F:1-37/Domain: link protein repeat homology (fragment) <LNKL>

F:58-139/Domain: link protein repeat homology <LNKL>

F:722-753/Domain: EGF homology <EG1>

F:760-791/Domain: EGF homology <EG2>

Query Match

Best Local Similarity 87.5%; Score 49; DB 2; Length 862;

Matches 8; Conservative 1; Mismatches 0; Gaps 0;

QY 1 CVIGYSGDXC 10

|| |||||:|

Db 744 CVPYSGDQC 753

#### RESULT 7

T14274

N:Alternate names: versican precursor, splice form V2 - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 05-May-2000

C:Accession: T14274

R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.

J. Biol. Chem. 273, 15758-15764, 1998

A:Title: Versican V2 is a major extracellular matrix component of the mature bovine brain

A:Reference number: 217954; MUID:98288320; PMID:9624174

A:Accession: T14274

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1643 <SCH>

A:Cross-references: EMBL:AF060458; NID:g32533303; PID:g32533304; PIDN:AAC24360.1

A:Experimental source: brain

C:Keywords: glycoprotein

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-1643/Product: versican, splice form V2 #status predicted <MAT>

F:57-331,352,817,965,1017,1333,1616,1626/Binding site: carbohydrate (Asn) (covalent) #st

Query Match

Best Local Similarity 87.5%; Score 49; DB 2; Length 1643;

Matches 8; Conservative 1; Mismatches 0; Gaps 0;

QY 1 CVIGYSGDXC 10

|| |||||:|

Db 1362 CVPYSGDRC 1371

#### RESULT 8

A60979

N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteogly

C:Species: Homo sapiens (man)

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Jan-2001

C:Accession: S06014; S43921; A60979; A30358; A29348; A45131; I54179

R:Zimmermann, D.R.; Ruoslahti, E.

EMBO J. 8, 2975-2981, 1989

A:Title: Multiple domains of the large fibroblast proteoglycan, versican.

A:Reference number: S06014; MUID:90059882; PMID:2583089

A:Accession: S06014

A:Molecule type: mRNA

A:Residues: 1-2409 <ZIM>

A:Cross-references: GB:X15998; NID:g37662; PIDN:CAA34128.1; PID:g37663

R:Yao, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.

Matrix Biol. 14, 213-225, 1994

A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cell

A:Reference number: S43921; MUID:95005762; PMID:7921538

A:Accession: S43921

A:Molecule type: mRNA

A:Residues: 208-440;1094-1385;1910-2246 <YAO>

R:Bignami, A.; Lane, W.S.; Andrews, D.; Dahl, D.

Brain Res. Bull. 22, 67-70, 1989

A:Title: Structural similarity of hyaluronate binding proteins in brain and cartilag

A:Reference number: A60979; MUID:89229983; PMID:2469524

A:Accession: A60979

A:Molecule type: protein

A:Residues: 171-210;289-303 <BIG>

R:Perides, G.; Lane, W.S.; Andrews, D.; Dahl, D.; Bignami, A.

J. Biol. Chem. 264, 5981-5987, 1989

A:Title: Isolation and partial characterization of a glial hyaluronate-binding prote

A:Reference number: A30358; MUID:89174663; PMID:2466833

A:Accession: A30358

A:Molecule type: protein

A:Residues: 24-50;80-87,'D',89-119;128-155;167-218;229-259,'IR';261-268;277-283,'G',

R:Krusius, T.; Gehlsen, K.R.; Ruoslahti, E.

J. Biol. Chem. 262, 13120-13125, 1987

A:Title: A fibroblast chondroitin sulfate proteoglycan core protein contains lectin-

A:Reference number: A29348; MUID:88007514; PMID:2820964

A:Accession: A29348

A:Molecule type: mRNA

A:Residues: 1725,'V',1727-2409 <KRU>

A:Cross-references: GB:J02814

R:Perides, G.; Rahemtulla, F.; Lane, W.S.; Asher, R.A.; Bignami, A.

J. Biol. Chem. 267, 23883-23887, 1992

A:Title: Isolation of a large aggregating proteoglycan from human brain.

A:Reference number: A45131; MUID:93054750; PMID:1429726

A:Contents: brain

A:Accession: A45131

A:Molecule type: protein

A:Residues: 21-22,'X',24-37 <PE2>

A:Experimental source: brain

A>Note: sequence extracted from NCBI backbone (NCBIP:118884)

R:Iozzo, R.V.; Naso, M.F.; Cannizzaro, L.A.; Wasmuth, J.J.; McPherson, J.D.

Genomics 14, 845-851, 1992

A:Title: Mapping of the versican proteoglycan gene (CSPG2) to the long arm of human

A:Reference number: I54179; MUID:93122792; PMID:1478664

A:Accession: I54179

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 251-347 <RES>

ating the differentiation of MDA-MB-453 cells.  
C/Superfamily: human ErbB kinase activator alpha, brain and thymus; EGF homology  
F:1-25/Domain: EGF homology (fragment) <EGF>

Query Match 82.1%; Score 46; DB 2; Length 57;  
Best Local Similarity 60.0%; Pred. No. 0.34;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CVIGYSGDXC 10  
I :|||:  
Db 16 CPVGTGDRC 25

RESULT 11

EGRT  
epidermal growth factor precursor - rat  
N/Alternate names: urogastrone precursor  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 31-Dec-1998 #sequence revision 14-Aug-1998 #text\_change 18-Jun-1999  
C/Accession: I52995; S05074; S01974; A25425; S18419; S08288  
R/Saggi, S.J.; Safirstein, R.; Price, P.M.  
DNA Cell Biol. 11, 481-487, 1992  
A/Title: Cloning and Sequencing of the Rat Preproepidermal Growth Factor cDNA: Comparison of the Rat and Human Genes  
A/Reference number: I52995; MUID:92398779; PMID:1524680  
A/Accession: I52995  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-1133 <RES>  
A/Cross-references: EMBL:U04842; NID:g440236; PIDN:AAB60436.1; PID:g440237  
R/Simpson, R.  
submitted to the EMBL Data Library, August 1988  
A/Reference number: S05074  
A/Accession: S05074  
A/Molecule type: mRNA  
A/Residues: 'W' 966 'RWL' 970-1023, 'NW' 1026-1108, 'SGAGVSGPQPFVVLE', 1126, 'HQ' <SWM>  
A/Cross-references: EMBL:X12748  
R/Dorow, D.S.; Simpson, R.J.  
Nucleic Acids Res. 16, 9338, 1988  
A/Title: Cloning and sequence analysis of a cDNA for rat epidermal growth factor.  
A/Reference number: S01974; MUID:89016634; PMID:3262867  
A/Accession: S01974  
A/Molecule type: mRNA  
A/Residues: 'W' 966 'RWL' 970-1023, 'NW' 1026-1108 <DOR>  
A/Cross-references: EMBL:X12748  
R/Simpson, R.J.; Smith, J.A.; Moritz, R.L.; O'Hare, M.J.; Rudland, P.S.; Morrison, J.  
Eur. J. Biochem. 153, 629-637, 1985  
A/Title: Rat epidermal growth factor: complete amino acid sequence.  
A/Reference number: A25425; MUID:86081810; PMID:3000782  
A/Accession: A25425  
A/Molecule type: protein  
A/Residues: 974-1021 <SIZ>  
R/R.Nishi, N.; Shimizu, C.; Okutani, T.; Kagawa, Y.; Takasuga, H.; Suno, M.; Wada, F.  
Biochim. Biophys. Acta 1095, 268-275, 1991  
A/Title: Rat prostatic growth factors: purification and characterization of high and low molecular weight forms  
A/Reference number: S18419; MUID:92069070; PMID:1958699  
A/Accession: S18419  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 974-1021 <NIS>  
R/R.Nexo, E.; Jorgensen, P.E.; Thim, L.; Roepstorff, P.  
Biochim. Biophys. Acta 1037, 388-393, 1990  
A/Title: Purification and characterization of a low and a high molecular weight form of rat epidermal growth factor  
A/Reference number: S08288; MUID:90181442; PMID:2310752  
A/Accession: S08288  
A/Molecule type: protein  
A/Residues: 974-1024 <NEX>  
C/Comment: Epidermal growth factor (EGF) stimulates the proliferation and differentiation of many types of animal cells.  
C/Comment: The EGF precursor is released in the pancreas, small intestine, mammary gland, and (in some species) placenta.  
C/Comment: Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor WTDF-can-  
F:1-25/Domain: signal sequence #status predicted <SIG>

F:22-1133/Product: epidermal growth factor proprotein, membrane-bound form #status predicted  
 F:22-1035/Region: EGF precursor long repeat <EXT>  
 F:44-480/Domain: EGF precursor long repeat <LR1>  
 F:47-86/Domain: LDL receptor YWTD-containing repeat homology <YW01>  
 F:87-128/Domain: LDL receptor YWTD-containing repeat homology <YW02>  
 F:129-170/Domain: LDL receptor YWTD-containing repeat homology <YW03>  
 F:171-212/Domain: LDL receptor YWTD-containing repeat homology <YW04>  
 F:213-257/Domain: LDL receptor YWTD-containing repeat homology <YW05>  
 F:258-302/Domain: LDL receptor YWTD-containing repeat homology <YW06>  
 F:319-355/Domain: EGF homology #status atypical <EG1>  
 F:361-396/Domain: EGF homology <EG2>  
 F:402-437/Domain: EGF homology <EG3>  
 F:440-477/Domain: EGF homology <EG4>  
 F:482-958/Region: EGF precursor long repeat <LR2>  
 F:485-525/Domain: LDL receptor YWTD-containing repeat homology <YW07>  
 F:526-568/Domain: LDL receptor YWTD-containing repeat homology <YW08>  
 F:569-611/Domain: LDL receptor YWTD-containing repeat homology <YW09>  
 F:612-655/Domain: LDL receptor YWTD-containing repeat homology <YW10>  
 F:656-696/Domain: LDL receptor YWTD-containing repeat homology <YW11>  
 F:697-739/Domain: LDL receptor YWTD-containing repeat homology <YW12>  
 F:747-782/Domain: EGF homology <EG5>  
 F:839-872/Domain: EGF homology <EG6>  
 F:878-914/Domain: EGF homology <EG7>  
 F:920-955/Domain: EGF homology <EG8>  
 F:974-1024/Product: epidermal growth factor #status experimental <NAT>  
 F:979-1015/Domain: EGF homology <EG9>  
 F:1036-1060/Domain: transmembrane #status predicted <TMM>  
 F:1061-1133/Domain: intracellular #status predicted <INT>  
 F:342-355,361-372,368-381,383-396,402-413,409-422,424-437,440-452,448-462,464-477,747-75  
 -1015/Disulfide bonds: #status predicted

Query Match 82.1%; Score 46; DB 1; Length 1133;

Best Local Similarity 70.0%; Pred. No. 6;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10

||||| :|:

DB 1006 CVIGYIGERC 1015

RESULT 12

T42218

slit-1 protein homolog - rat

N:Alternate names: MEGF4 protein

C:Species: Rattus norvegicus (Norway rat)

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Aug-2002

C:Accession: T42218

R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomics 51, 27-34, 1998

A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs

A:Reference number: Z14126; MUID:98360089; PMID:9693030

A:Accession: T42218

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1531 <NAK>

A:Cross-references: EMBL:AB011530; NID:g3449289; PIDN:BAA32460.1; PID:g3449290

A:Experimental source: strain Sprague-Dawley; brain

C:Genetics:

A:Gene: MEGF4

C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein

Query Match

Best Local Similarity 82.1%; Score 46; DB 2; Length 1531;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10

||||| :|:

DB 1071 CVPGTGDC 1080

RESULT 13

A31246

neurogenic protein Delta precursor - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster  
 C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 02-Aug-2002  
 C:Accession: A31246  
 R:Kopczynski, C.C.; Alton, A.K.; Rechtel, K.; Kooh, P.J.; Muskavitch, M.A.T.  
 Genes Dev. 2, 1733-1735, 1988  
 A:Title: Delta, a Drosophila neurogenic gene, is transcriptionally complex and encoded  
 A:Reference number: A31246; MUID:89196890; PMID:3149249  
 A:Accession: A31246  
 A:Molecule type: mRNA  
 A:Residues: 1-832 <KOP>  
 A:Cross-references: GB:Y00222  
 C:Genetics:

A:Gene: FlyBase:DL

A:Cross-references: FlyBase:FBgn0000463

F:295-328/Domain: EGF homology <EGX1>

F:422-450/Domain: EGF homology <EGF1>

F:457-488/Domain: EGF homology <EGF>

F:533-564/Domain: EGF homology <EGF3>

Query Match 80.4%; Score 45; DB 2; Length 832;

Best Local Similarity 70.0%; Pred. No. 6.6;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10

||||| :|:

DB 319 CAPGYSGDDC 328

RESULT 14

SI9087

gene Delta protein precursor - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 02-Aug-2002

C:Accession: SI9087

R:Muskavitch, M.A.T.

A:Reference number: SI9087

A:Accession: SI9087

A:Molecule type: mRNA

A:Residues: 1-833 <MUS>

A:Cross-references: EMBL:Y00222

C:Genetics:

A:Gene: FlyBase:DL

A:Cross-references: FlyBase:FBgn0000463

C:Superfamily: neurogenic protein delta; EGF homology

F:335-371/Domain: EGF homology <EGF1>

F:378-415/Domain: EGF homology <EGX1>

F:457-488/Domain: EGF homology <EGF>

F:533-564/Domain: EGF homology <EGF3>

Query Match 80.4%; Score 45; DB 2; Length 833;

Best Local Similarity 70.0%; Pred. No. 6.6;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10

||||| :|:

DB 319 CAPGYSGDDC 328

RESULT 15

S00670

neurogenic repetitive locus delta protein precursor - fruit fly (Drosophila melanogaster)

N:Alternate names: gene DL protein

C:Species: Drosophila melanogaster

C>Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 02-Aug-2002

C:Accession: S00670; A26637

R:Vaessin, H.; Bremer, K.A.; Knust, E.; Campos-Ortega, J.A.

EMBO J. 6, 3431-3440, 1987

A:Title: The neurogenic gene Delta of Drosophila melanogaster is expressed in neurog

A:Reference number: S00670

A:Accession: S00670

A:Molecule type: mRNA  
A:Residues: 1-880 <VAE>  
A:Cross-references: EMBL:X06289; NID:g7852; PID:g7853  
R:Knust, E.; Dietrich, U.; Tepass, U.; Bremer, K.A.; Weigelt, D.; Vaessin, H.; Campos-Ortega J. 6. 761-766, 1987  
A:Title: EGF homologous sequences encoded in the genome of Drosophila melanogaster, and  
A:Reference number: A91081; MUID:87218537; PMID:3107986  
A:Accession: A26637  
A:Molecule type: mRNA  
A:Residues: 422-436, 'ET', 439-458, 'A', 460-489, 'T', 491-621 <KNU>  
A:Cross-references: GB:X05140; NID:g7851; PIDN:CAA28786.1; PID:g929563  
C:Genetics:  
A:Gene: Delta; DL  
A:Cross-references: FlyBase:FBgn0000463  
C:Superfamily: neurogenic protein delta; EGF homology  
C:Keywords: transmembrane protein  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-88/Product: neurogenic repetitive locus delta protein #status predicted <MAT>  
F:457-488/Domain: EGF homology <EGF1>  
F:533-564/Domain: EGF homology <EGF2>

Query Match 80.4%; Score 45; DB 2; Length 880;  
Best Local Similarity 70.0%; Pred. No. 7;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CVIGYSGDXC 10  
Db 319 CAPGYSGDDC 328

Search completed: July 2, 2003, 07:26:26  
Job time : 12 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:02 ; Search time 10.3333 Seconds  
(without alignments)  
40.138 Million cell updates/sec

Title: US-09-673-785D-9  
Perfect score: 56  
Sequence: 1 CVIGYSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	1217	1	EGF_MOUSE
2	51	91.1	2139	1	CRB_DROME
3	49	87.5	862	1	PGCV_MACNE
4	49	87.5	3381	1	PGCV_BOVIN
5	49	87.5	3396	1	PGCV_HUMAN
6	48	85.7	2920	1	CLR2_MOUSE
7	46	82.1	756	1	NRG2_MOUSE
8	46	82.1	1133	1	EGF_RAT
9	45	80.4	833	1	DL_DROME
10	45	80.4	883	1	PGCB_RAT
11	45	80.4	1207	1	EGF_HUMAN
12	45	80.4	2319	1	NTC3_RAT
13	45	80.4	2738	1	PGCV_RAT
14	45	80.4	3358	1	PGCV_MOUSE
15	44	78.6	1064	1	FBP1_STRPU
16	44	78.6	2470	1	NTC2_MOUSE
17	44	78.6	2471	1	NTC2_HUMAN
18	44	78.6	2471	1	NTC2_RAT
19	44	78.6	2476	1	ZAN_PIG
20	44	78.6	4289	1	TENX_HUMAN
21	43	76.8	2321	1	NTC3_HUMAN
22	43	76.8	3672	1	LM22_CAEEL
23	42	75.0	53	1	EGF_PIG
24	42	75.0	57	1	Y402_ARCFU
25	42	75.0	473	1	FP2_MVTGA
26	42	75.0	883	1	PGCB_MOUSE
27	42	75.0	912	1	PGCB_BOVIN
28	42	75.0	1408	1	SERR_DROME
29	42	75.0	2318	1	NTC3_MOUSE
30	41	73.2	597	1	BP10_PARLI
31	41	73.2	1049	1	ADP1_YEAST
32	41	73.2	1192	1	LMG2_MOUSE
33	41	73.2	2437	1	NTC1_BRARE

34	40	71.4	427	1	MFGM_RAT	P70490	rattus norv
35	40	71.4	463	1	MFGM_MOUSE	P21956	mus musculus
36	40	71.4	723	1	DL11_HUMAN	O00548	homo sapien
37	40	71.4	1213	1	JAG3_BRARE	Q90y54	brachydanio
38	40	71.4	1218	1	JAG1_HUMAN	P78504	homo sapien
39	40	71.4	1218	1	JAG1_MOUSE	Q9qxx0	mus musculus
40	40	71.4	1219	1	JAG1_RAT	Q63722	rattus norv
41	40	71.4	1242	1	JAG1_BRARE	Q90y57	brachydanio
42	40	71.4	1429	1	LI12_CAEEL	P14585	caenorhabdi
43	40	71.4	2531	1	NTC1_MOUSE	Q01705	mus musculus
44	40	71.4	2531	1	NTC1_RAT	Q07008	rattus norv
45	40	71.4	2703	1	NOTC_DROME	P07077	drosophila

ALIGNMENTS

RESULT 1  
ID EGF\_MOUSE STANDARD; PRT; 1217 AA.  
AC P01132;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal growth factor].  
DE EGF.  
GN EGF.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83223630; PubMed=6602382;  
RA Scott J., Urdea M., Quiroga M., Sanchez-Pescador R., Fong N.M., Selby M., Rutter W.J., Bell G.I.;  
RT "Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins.";  
RL Science 221:236-240(1983).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83219309; PubMed=6304537;  
RA Gray A., Dull T.J., Ullrich A.;  
RT "Nucleotide sequence of epidermal growth factor cDNA predicts a Nature 303:722-725(1983).  
RN [3]  
RP SEQUENCE OF 977-1029.  
RX MEDLINE=73048516; PubMed=4636327;  
RA Savage C.R. Jr., Inagami T., Cohen S.;  
RT "The primary structure of epidermal growth factor.";  
RN [4]  
RP J. Biol. Chem. 247:7612-7621(1972).  
DISULFIDE BONDS.  
RX MEDLINE=74025498; PubMed=4750422;  
RA Savage C.R. Jr., Hash J.H., Cohen S.;  
RT "Epidermal growth factor. Location of disulfide bonds.";  
RN [5]  
RP J. Biol. Chem. 248:7669-7672(1973).  
STRUCTURE BY NMR OF 977-1029.  
RX MEDLINE=92118798; PubMed=1731873;  
RA Montellone G.T., Wuethrich K., Burgess A.W., Nice E.C., Wagner G., Gibson K.D., Scheraga H.A.;  
RT "Solution structure of murine epidermal growth factor determined by NMR spectroscopy and refined by energy minimization with restraints.";  
RN [6]  
RP Biochemistry 31:236-249(1992).  
STRUCTURE BY NMR OF 977-1029.  
RX MEDLINE=93075811; PubMed=1445923;  
RA Kohda D., Inagaki F.;  
RT "Three-dimensional nuclear magnetic resonance structures of mouse epidermal growth factor in acidic and physiological pH solutions.";  
RN

FT	DISULFID	414	427		BY SIMILARITY.					
FT	DISULFID	429	442		BY SIMILARITY.					
FT	DISULFID	445	457		BY SIMILARITY.					
FT	DISULFID	453	467		BY SIMILARITY.					
FT	DISULFID	469	482		BY SIMILARITY.					
FT	DISULFID	751	762		BY SIMILARITY.					
FT	DISULFID	758	771		BY SIMILARITY.					
FT	DISULFID	773	786		BY SIMILARITY.					
FT	DISULFID	842	853		BY SIMILARITY.					
FT	DISULFID	847	862		BY SIMILARITY.					
FT	DISULFID	864	875		BY SIMILARITY.					
FT	DISULFID	881	895		BY SIMILARITY.					
FT	DISULFID	888	904		BY SIMILARITY.					
FT	DISULFID	906	917		BY SIMILARITY.					
FT	DISULFID	923	936		BY SIMILARITY.					
FT	DISULFID	930	945		BY SIMILARITY.					
FT	DISULFID	947	958		BY SIMILARITY.					
FT	DISULFID	982	996		BY SIMILARITY.					
FT	DISULFID	990	1007		BY SIMILARITY.					
FT	DISULFID	1009	1018		BY SIMILARITY.					
FT	DOMAIN	1024	1029		BY SIMILARITY.					
FT										
FT	CARBOHYD	111	111		N-LINKED (GLCNAC. .) (POTENTIAL).					
FT	CARBOHYD	410	410		N-LINKED (GLCNAC. .) (POTENTIAL).					
FT	CARBOHYD	810	810		N-LINKED (GLCNAC. .) (POTENTIAL).					
FT	CARBOHYD	944	944		N-LINKED (GLCNAC. .) (POTENTIAL).					
FT	CONFLICT	790	790		D -> Y (IN REF. 2).					
FT	CONFLICT	1048	1048		A -> S (IN REF. 2).					
FT	STRAND	995	997							
FT	STRAND	1006	1008							
FT	STRAND	1010	1010							
FT	TURN	1011	1012							
FT	STRAND	1013	1014							
FT	STRAND	1020	1021							
SQ	SEQUENCE	1217 AA;	133144 MW;	A9C7F3D512F82873 CRC64;						
Query Match 100.0%; Score 56; DB 1; Length 1217;										
Best Local Similarity 90.0%; Pred. No. 0.027;										
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps										
QY	1 CVIGYSGDXC 10									
	::									
Db	1009 CVIGYSGDCR 1018									
RESULT 2										
CRB_DROME										
ID	CRB_DROME	STANDARD;	PRT;	2139 AA.						
AC	F10040;									
DT	01-MAR-1989 (Rel. 10, Created)									
DT	01-MAY-1991 (Rel. 18, Last sequence update)									
DT	15-JUN-2002 (Rel. 41, Last annotation update)									
DE	Crumbs protein precursor (95F).									
GN	CRB.									
OS	Drosophila melanogaster (Fruit fly).									
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;									
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;									
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.									
NCBI_TaxID=7227;										
[1]	SEQUENCE FROM N.A.									
RN	STRAIN-Oregon-R; TISSUE=Embryo;									
RC	MEDLINE=90263104; PubMed=2344615;									
RA	Tepass U., Theres C., Knust E.;									
RT	"Crumbs encodes an EGF-like protein expressed on apical membranes of									
RT	drosophila epithelial cells and required for organization of									
RL	epithella.";									
RN	Cell 61:787-799(1990).									
RP	SEQUENCE OF 1663-1955 FROM N.A.									
RC	TISSUE=Embryo;									
RA	MEDLINE=87218537; PubMed=3107986;									
RX	Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.;									

RA Vaessin H., Campos-Ortega J.A.;  
 RT "EGF homologous sequences encoded in the genome of Drosophila  
 RL melanogaster, and their relation to neurogenic genes.";   
 RL ENBO J. 6:761-766(1987).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,  
 CC POLARITY. IT MAY ACT AS A SIGNAL.  
 CC POSSIBLE FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL  
 CC SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).  
 CC -1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M33753; AAA28428.1; ALT\_SEQ.  
 DR EMBL; X05144; CAA28793.1;  
 DR PIR; B26637; B26637.  
 DR PIR; A35672; A35672.  
 DR HSSP; P00740; IEDM.  
 DR FlyBase; FBgn0000368; crb.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR001438; EGF\_II.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR Pfam; PF00008; EGF; 26.  
 DR Pfam; PF00054; laminin\_G; 3.  
 DR PRINTS; PR00010; EGFBL00D.  
 DR SMART; SM00179; EGF\_CA; 11.  
 DR SMART; SM00001; EGF\_like; 16.  
 DR SMART; SM00282; LamG; 3.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 15.  
 DR PROSITE; PS00022; EGF\_1; 26.  
 DR PROSITE; PS01186; EGF\_2; 17.  
 DR PROSITE; PS01187; EGF\_CA; 12.  
 DR PROSITE; PS00025; LAM\_G\_DOMAIN; 3.  
 KW Differentiation; Repeat; EGF-like domain; Transmembrane;  
 KW Glycoprotein; Signal; Phosphorylation.  
 FT SIGNAL 1 90  
 FT CHAIN 91 2139 CRUMBS PROTEIN.  
 FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 2085 2111 POTENTIAL.  
 FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 267 303 EGF-LIKE 1.  
 FT DOMAIN 306 343 EGF-LIKE 2.  
 FT DOMAIN 348 386 EGF-LIKE 3.  
 FT DOMAIN 388 425 EGF-LIKE 4.  
 FT DOMAIN 427 463 EGF-LIKE 5.  
 FT DOMAIN 464 500 EGF-LIKE 6.  
 FT DOMAIN 501 532 EGF-LIKE 7.  
 FT DOMAIN 545 581 EGF-LIKE 8.  
 FT DOMAIN 582 611 EGF-LIKE 9.  
 FT DOMAIN 609 646 EGF-LIKE 10.  
 FT DOMAIN 648 685 EGF-LIKE 11. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 687 723 EGF-LIKE 12. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 725 761 EGF-LIKE 13. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 763 800 EGF-LIKE 14. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 802 838 EGF-LIKE 15. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 840 902 EGF-LIKE 16.  
 FT DOMAIN 904 940 EGF-LIKE 17. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 942 978 EGF-LIKE 18. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 980 1021 EGF-LIKE 19.  
 FT DOMAIN 1023 1205 LAMININ G-LIKE 1.  
 FT DOMAIN 1207 1243 EGF-LIKE 20.  
 FT DOMAIN 1250 1480 LAMININ G-LIKE 2.  
 FT DOMAIN 1481 1517 EGF-LIKE 21.

FT DOMAIN 1558  
 FT DOMAIN 1759 LAMININ G-LIKE 3.  
 FT DOMAIN 1795 EGF-LIKE 22.  
 FT DOMAIN 1833 EGF-LIKE 23. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 1835 EGF-LIKE 24. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 1874 EGF-LIKE 25.  
 FT DOMAIN 1915 EGF-LIKE 26.  
 FT DOMAIN 1951 EGF-LIKE 27. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 1953 EGF-LIKE 28. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 1991 EGF-LIKE 29.  
 FT DOMAIN 2029 BY SIMILARITY.  
 FT DOMAIN 2030 BY SIMILARITY.  
 FT DISULFID 271 282 BY SIMILARITY.  
 FT DISULFID 276 291 BY SIMILARITY.  
 FT DISULFID 293 302 BY SIMILARITY.  
 FT DISULFID 310 321 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 333 342 BY SIMILARITY.  
 FT DISULFID 352 363 BY SIMILARITY.  
 FT DISULFID 357 374 BY SIMILARITY.  
 FT DISULFID 376 385 BY SIMILARITY.  
 FT DISULFID 392 403 BY SIMILARITY.  
 FT DISULFID 397 412 BY SIMILARITY.  
 FT DISULFID 414 424 BY SIMILARITY.  
 FT DISULFID 431 442 BY SIMILARITY.  
 FT DISULFID 436 451 BY SIMILARITY.  
 FT DISULFID 453 462 BY SIMILARITY.  
 FT DISULFID 468 479 BY SIMILARITY.  
 FT DISULFID 473 488 BY SIMILARITY.  
 FT DISULFID 490 499 BY SIMILARITY.  
 FT DISULFID 505 515 BY SIMILARITY.  
 FT DISULFID 509 520 BY SIMILARITY.  
 FT DISULFID 522 531 BY SIMILARITY.  
 FT DISULFID 549 562 BY SIMILARITY.  
 FT DISULFID 556 569 BY SIMILARITY.  
 FT DISULFID 571 580 BY SIMILARITY.  
 FT DISULFID 586 597 BY SIMILARITY.  
 FT DISULFID 591 602 BY SIMILARITY.  
 FT DISULFID 604 610 BY SIMILARITY.  
 FT DISULFID 613 624 BY SIMILARITY.  
 FT DISULFID 618 634 BY SIMILARITY.  
 FT DISULFID 636 645 BY SIMILARITY.  
 FT DISULFID 652 664 BY SIMILARITY.  
 FT DISULFID 659 673 BY SIMILARITY.  
 FT DISULFID 675 684 BY SIMILARITY.  
 FT DISULFID 691 702 BY SIMILARITY.  
 FT DISULFID 696 711 BY SIMILARITY.  
 FT DISULFID 713 722 BY SIMILARITY.  
 FT DISULFID 729 740 BY SIMILARITY.  
 FT DISULFID 734 749 BY SIMILARITY.  
 FT DISULFID 751 760 BY SIMILARITY.  
 FT DISULFID 767 778 BY SIMILARITY.  
 FT DISULFID 772 787 BY SIMILARITY.  
 FT DISULFID 789 799 BY SIMILARITY.  
 FT DISULFID 806 817 BY SIMILARITY.  
 FT DISULFID 811 826 BY SIMILARITY.  
 FT DISULFID 828 837 BY SIMILARITY.  
 FT DISULFID 844 855 BY SIMILARITY.  
 FT DISULFID 849 890 BY SIMILARITY.  
 FT DISULFID 892 901 BY SIMILARITY.  
 FT DISULFID 908 919 BY SIMILARITY.  
 FT DISULFID 913 928 BY SIMILARITY.  
 FT DISULFID 930 939 BY SIMILARITY.  
 FT DISULFID 946 957 BY SIMILARITY.  
 FT DISULFID 952 966 BY SIMILARITY.  
 FT DISULFID 968 977 BY SIMILARITY.  
 FT DISULFID 984 995 BY SIMILARITY.  
 FT DISULFID 989 1009 BY SIMILARITY.  
 FT DISULFID 1011 1020 BY SIMILARITY.  
 FT DISULFID 1211 1222 BY SIMILARITY.  
 FT DISULFID 1216 1231 BY SIMILARITY.  
 FT DISULFID 1233 1242 BY SIMILARITY.  
 FT DISULFID 1243 1296 BY SIMILARITY.  
 FT DISULFID 1485 1505 BY SIMILARITY.  
 FT DISULFID 1490 1516 BY SIMILARITY.  
 FT DISULFID 1507 1516 BY SIMILARITY.  
 FT DISULFID 1763 1774 BY SIMILARITY.

```

FT DISULFID 1768 1783 BY SIMILARITY.
FT DISULFID 1785 1794 BY SIMILARITY.
FT DISULFID 1801 1812 BY SIMILARITY.
FT DISULFID 1806 1821 BY SIMILARITY.
FT DISULFID 1823 1832 BY SIMILARITY.
FT DISULFID 1839 1850 BY SIMILARITY.
FT DISULFID 1844 1859 BY SIMILARITY.
FT DISULFID 1861 1870 BY SIMILARITY.
FT DISULFID 1878 1889 BY SIMILARITY.
FT DISULFID 1883 1903 BY SIMILARITY.
FT DISULFID 1905 1914 BY SIMILARITY.
FT DISULFID 1919 1930 BY SIMILARITY.
FT DISULFID 1924 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1957 1968 BY SIMILARITY.
FT DISULFID 1962 1977 BY SIMILARITY.
FT DISULFID 1979 1988 BY SIMILARITY.
FT DISULFID 1995 2008 BY SIMILARITY.
FT DISULFID 2002 2017 BY SIMILARITY.
FT DISULFID 2019 2028 BY SIMILARITY.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 91.1%; Score 51; DB 1; Length 2139;
Best Local Similarity 70.0%; Pred. No. 0.37;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10
Db 571 CAVIGSGDRC 580
1:|||||:1

RESULT 3
PGCV_MACNE STANDARD; PRT; 862 AA.
AC Q28858: Q28859; Q28860;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Versican core protein (Large fibroblast proteoglycan) (Chondroitin
DE sulfate proteoglycan core protein 2) (Fragments).
GN CSPG2.
OS Macaca nemestrina (Pig-tailed macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9545;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Aortic smooth muscle;
RX MEDLINE=95005762; PubMed=7921538;
RA Yao L.Y., Moody C., Schoenher E., Wight T.N., Sandell L.J.;
RT "Identification of the proteoglycan versican in aorta and smooth
RT muscle cells by DNA sequence analysis, in situ hybridization and
RT immunohistochemistry."
RL Matrix Biol. 14:213-225(1994).
CC -1- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronan.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC similarity).
CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (By similarity).
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.

```

```

CC CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; S72412; AAA65593.2; -
CC DR EMBL; S72413; AAA65594.2; -
CC DR EMBL; S72414; AAA65595.2; -
CC DR HSP; P01132; IEGF.
CC DR InterPro; IPR000152; Asx_hydroxyl.
CC DR InterPro; IPR000561; EGF-like.
CC DR InterPro; IPR000742; EGF_2.
CC DR InterPro; IPR001881; EGF_Ca.
CC DR InterPro; IPR001438; EGF_II.
CC DR InterPro; IPR001304; Lectin_C.
CC DR InterPro; IPR000538; Link.
CC DR PRINTS; PR00010; EGFBL00D.
CC DR ProDom; PD000918; Link; 2.
CC DR SMART; SM00034; CLECT; 1.
CC DR SMART; SM00179; EGF_CA; 1.
CC DR SMART; SM00001; EGF_Like; 1.
CC DR SMART; SM00445; LINK; 1.
CC DR PROSITE; PS00010; ASX_HYDROXYL; 1.
CC DR PROSITE; PS00022; EGF_1; 2.
CC DR PROSITE; PS01186; EGF_2; 1.
CC DR PROSITE; PS01187; EGF_CA; 1.
CC DR PROSITE; PS01241; LINK; 1.
CC DR PROSITE; PS00615; C-TYPE-LECTIN_1; PARTIAL.
CC DR PROSITE; PS50041; C-TYPE-LECTIN_2; 1.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Repeat;
KW EGF-like domain; Calcium.
FT NON_TER 1 1
FT DOMAIN <1 37 LINK 1.
FT DOMAIN 58 139 LINK 2.
FT DOMAIN 141 >233 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
FT DOMAIN)
FT NON_CONS 233 234 GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN,
FT DOMAIN <234 >525 SIMILAR TO CHONDROITIN SULFATE ATTACHMENT
FT SITE IN COLLAGEN TYPE IX (BY SIMILARITY).
FT NON_CONS 525 526 EGF-LIKE 1.
FT DOMAIN 718 754 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 756 792 EGF-LIKE 1.
FT DISULFID 63 138 C-TYPE LECTIN.
FT DISULFID 87 108 BY SIMILARITY.
FT DISULFID 722 733 BY SIMILARITY.
FT DISULFID 727 742 BY SIMILARITY.
FT DISULFID 744 753 BY SIMILARITY.
FT DISULFID 760 771 BY SIMILARITY.
FT DISULFID 765 780 BY SIMILARITY.
FT DISULFID 782 791 BY SIMILARITY.
FT CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 563 563 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 696 696 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 862 862
SQ SEQUENCE 862 AA; 95583 MW; ASD5F6153A74BB39 CRC64;

Query Match 87.5%; Score 49; DB 1; Length 862;
Best Local Similarity 80.0%; Pred. No. 0.35;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10

```



```

FT CARBOHYD 2074 2074 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2103 2103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2263 2263 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2290 2290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2356 2356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2623 2623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2641 2641 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2919 2919 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3052 3052 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3354 3354 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3364 3364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBSPLIC 349 349 P -> R (IN ISOFORM V1 AND V3).
FT VARSPLIC 350 1336 MISSING (IN ISOFORM V1).
FT VARSPLIC 1337 3074 MISSING (IN ISOFORM V2).
FT VARSPLIC 350 3074 MISSING (IN ISOFORM V3).
FT CONFLICT 25 25 MISSING (IN REF. 2).
FT CONFLICT 51 51 N -> D (IN REF. 2).
FT CONFLICT 89 89 Q -> D (IN REF. 2).
FT CONFLICT 96 96 C -> R (IN REF. 2).
FT CONFLICT 346 346 C -> R (IN REF. 2).
SQ SEQUENCE 3381 AA; 369984 MW; F09716FA7778D459 CRC64;

Query Match 87.5%; Score 49; DB 1; Length 3381;
Best Local Similarity 80.0%; Pred. No. 1.3;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
DB 3100 CVPGYSGDRC 3109

RESULT 5
PGCV HUMAN STANDARD; PRT: 3396 AA.
AC P13611; P20754; G9UNW5; Q13010; Q13189; Q15123;
AD 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP).
GN CSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM V0).
RX MEDLINE=95105188; PubMed=7528742;
RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
RT "Characterization of the complete genomic structure of the human
RT versican gene and functional analysis of its promoter.";
RL J. Biol. Chem. 269:32999-33008(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM V1).
RX TISSUE=Placenta;
RX MEDLINE=90059882; PubMed=2583089;
RA Zimmermann D.R., Ruoslahti E.;
RT "Multiple domains of the large fibroblast proteoglycan, versican.";
RL EMBL J. 8:2975-2981(1989).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM V2).
RX TISSUE=Glial tumor;
RX MEDLINE=95105187; PubMed=7806529;
RA Dours-Zimmermann M.T., Zimmermann D.R.;
RT "A novel glycosaminoglycan attachment domain identified in two
RT alternative splice variants of human versican.";
RL J. Biol. Chem. 269:32992-32998(1994).
RN [4]
RP SEQUENCE OF 2711-3396 FROM N.A.
RX TISSUE=Lung fibroblast;
RX MEDLINE=88007514; PubMed=2820964;
RA Krusius T., Gehlsen K.R., Ruoslahti E.;

```

```

RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
RT lectin-like and growth factor-like sequences.";
RL J. Biol. Chem. 262:13120-13125(1987).
RN [5]
RP SEQUENCE OF 251-347 FROM N.A.
RX MEDLINE=93122792; PubMed=1478664;
RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
RA McPherson J.D.;
RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
RT human chromosome 5 (9q12-5q14).";
RL Genomics 14:845-851(1992).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM V3).
RX TISSUE=Brain;
RX MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues.";
RL J. Biol. Chem. 270:3914-3918(1995).
RN [7]
RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
RX TISSUE=Arctic smooth muscle;
RX MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wight T.N.;
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [8]
RP PARTIAL SEQUENCE.
RX TISSUE=Brain;
RX MEDLINE=89174663; PubMed=2466833;
RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
RT "Isolation and partial characterization of a glial
RT hyaluronate-binding protein.";
RL J. Biol. Chem. 264:5981-5987(1989).
RN [9]
RP TISSUE SPECIFICITY OF ISOFORMS.
RX MEDLINE=96213482; PubMed=8627343;
RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
RT "Differential expression of versican isoforms in brain tumors.";
RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
CC -1- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS: At least 5 isoforms; V0 (shown here), V1,
CC V2, V3 and Vint; are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
CC in normal brain, gliomas, medulloblastomas, schwannomas,
CC neurofibromas, and meningiomas; V2 is restricted to normal brain
CC and gliomas; V3 is found in all these tissues except
CC medulloblastomas.
CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; U16306; AAA65018.1; -
EMBL; X15998; CAA34128.1; -
EMBL; S52488; AAB24878.1; -

```

```
DR EMBL; U26555; AAA67565.1; -.
DR EMBL; D32039; BAA06801.1; -.
DR EMBL; J02814; AAA36437.1; -.
DR EMBL; AF084545; AAD48545.1; -.
DR PIR; S06014; S06014.
DR PIR; A29348; A29348.
DR PIR; A30358; A30358.
DR HSSP; P01132; IEGF.
DR Genew; HGNC:2464; CSPG2.
DR MIM; 118661; -.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_C; 1.
DR Pfam; PF00084; sushi; 1.
DR PRINTS; PR00193; xlink; 2.
DR PRODOM; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF-like; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE LECTIN_2; 1.
DR Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 3396
FT DOMAIN 37 137
FT DOMAIN 167 244
FT DOMAIN 265 346
FT DOMAIN 348 1335
FT DOMAIN 1336 3089
FT DOMAIN 3089 3125
FT DOMAIN 3127 3163
FT DOMAIN 3176 3290
FT DOMAIN 3295 3353
FT DISULFID 44 130
FT DISULFID 172 243
FT DISULFID 196 217
FT DISULFID 270 345
FT DISULFID 294 315
FT DISULFID 3093 3104
FT DISULFID 3098 3113
FT DISULFID 3115 3124
FT DISULFID 3131 3142
FT DISULFID 3136 3151
FT DISULFID 3153 3162
FT DISULFID 3169 3180
FT DISULFID 3197 3289
FT DISULFID 3265 3281
FT DISULFID 3296 3339
FT DISULFID 3325 3352
FT CARBOHYD 57 57
FT CARBOHYD 330 330
EMBL; U26555; AAA67565.1; -.
EMBL; D32039; BAA06801.1; -.
EMBL; J02814; AAA36437.1; -.
EMBL; AF084545; AAD48545.1; -.
PIR; S06014; S06014.
PIR; A29348; A29348.
PIR; A30358; A30358.
HSSP; P01132; IEGF.
Genew; HGNC:2464; CSPG2.
MIM; 118661; -.
InterPro; IPR000152; Asx_hydroxyl.
InterPro; IPR000561; EGF-like.
InterPro; IPR000742; EGF_2.
InterPro; IPR001881; EGF_Ca.
InterPro; IPR001438; EGF_II.
InterPro; IPR003599; Ig.
InterPro; IPR003006; Ig_MHC.
InterPro; IPR001304; Lectin_C.
InterPro; IPR000538; Link.
InterPro; IPR000436; Sushi_SCR_CCP.
Pfam; PF00008; EGF; 2.
Pfam; PF00047; Ig; 1.
Pfam; PF00059; lectin_C; 1.
Pfam; PF00084; sushi; 1.
PRINTS; PR00193; xlink; 2.
PRODOM; PD000918; Link; 2.
SMART; SM00032; CCP; 1.
SMART; SM00034; CLECT; 1.
SMART; SM00179; EGF_CA; 1.
SMART; SM00001; EGF-like; 1.
SMART; SM00409; IG; 1.
SMART; SM00445; LINK; 2.
PROSITE; PS00010; ASX_HYDROXYL; 1.
PROSITE; PS00022; EGF_1; 2.
PROSITE; PS01186; EGF_2; 1.
PROSITE; PS01187; EGF_CA; 1.
PROSITE; PS01241; LINK; 2.
PROSITE; PS00615; C-TYPE LECTIN_1; 1.
PROSITE; PS50041; C-TYPE LECTIN_2; 1.
Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
Hyaluronic acid; Alternative splicing.
SIGNAL 1 20
CHAIN 21 3396
DOMAIN 37 137
DOMAIN 167 244
DOMAIN 265 346
DOMAIN 348 1335
DOMAIN 1336 3089
DOMAIN 3089 3125
DOMAIN 3127 3163
DOMAIN 3176 3290
DOMAIN 3295 3353
DISULFID 44 130
DISULFID 172 243
DISULFID 196 217
DISULFID 270 345
DISULFID 294 315
DISULFID 3093 3104
DISULFID 3098 3113
DISULFID 3115 3124
DISULFID 3131 3142
DISULFID 3136 3151
DISULFID 3153 3162
DISULFID 3169 3180
DISULFID 3197 3289
DISULFID 3265 3281
DISULFID 3296 3339
DISULFID 3325 3352
CARBOHYD 57 57
CARBOHYD 330 330
FT CARBOHYD 615 615
FT CARBOHYD 782 782
FT CARBOHYD 809 809
FT CARBOHYD 1332 1332
FT CARBOHYD 1398 1398
FT CARBOHYD 1442 1442
FT CARBOHYD 1468 1468
FT CARBOHYD 1663 1663
FT CARBOHYD 1898 1898
FT CARBOHYD 2179 2179
FT CARBOHYD 2272 2272
FT CARBOHYD 2280 2280
FT CARBOHYD 2360 2360
FT CARBOHYD 2385 2385
Query Match 87.5%; Score 49; DB 1; Length 3396;
Best Local Similarity 80.0%; Pred.No. 1.3;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CVIGVSGDXC 10
Db 3115 CVPVSGDQC 3124
RESULT 6
CLR2_MOUSE STANDARD; PRT; 2920 AA.
ID Q9ROM0; Q9Z2R4; Q99K26;
AC DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Flamingo 1)
DE (mEmil).
GN CELSR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. PubMed=10490098;
RA MEDLINE=9418630; PubMed=10490098;
RA Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RA Takeichi M., Uemura T.;
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
RT polarity under the control of frizzled.";
RL Cell 98:585-595(1999).
RP [2]
RP SEQUENCE OF 1913-2796 FROM N.A., AND TISSUE SPECIFICITY.
RX PubMed=10790539;
RA Formstone C.J., Barclay J., Rees M., Little P.F.R.;
RT "Chromosomal localization of Celstr2 and Celstr3 in the mouse; Celstr3 is
RT a candidate for the tipy (tip) lethal mutant on chromosome 9.";
RL Mamm. Genome 11:392-394(2000).
RN [3]
RP SEQUENCE OF 2014-2920 FROM N.A.
RC TISSUE-Breast tumor;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP DEVELOPMENTAL STAGE.
RX PubMed=11850187;
RA Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;
RT "Developmental expression profiles of Celstr (Flamingo) genes in the
RT mouse.";
RL Mech. Dev. 112:157-160(2002).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- DEVELOPMENTAL STAGE: Predominantly expressed in the developing
CC CNS, the emerging dorsal root ganglia and cranial ganglia. In the
CC CNS, expression is uniform along the rostrocaudal axis. During
CC gastrulation, it is expressed within the anterior neural ectoderm.
CC At E10, expression is strong in the ventricular zones (Vz) in all
```

sectors of the brain, and lower in the marginal zones (MZ).  
 Between E12 and E15, expression is prominent in the brain. It is  
 strong in VZ, lower in MZ, except in telecephalic MZ where it is  
 predominant. The intensity is higher in all VZ, and lower in  
 differentiating fields than in VZ, except in the cerebral  
 hemispheres, and to a lesser extent in the tectum and cerebellum.  
 A weak expression is also observed in the fetal lungs, kidney and  
 epithelia. In the newborn and postnatal stages, expression remains  
 restricted to the VZ as well as in migrating and postmigratory  
 cells throughout the brain.

---  
 TISSUE SPECIFICITY: Expressed in the CNS and in the eye.  
 ---  
 -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 -1- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.  
 -1- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.  
 -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.  
 -1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAIN.  
 -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.  
 ---

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 EMBL; AB028499; BAA84070.1; -;  
 EMBL; AF031573; AAC68837.1; -;  
 EMBL; BC005499; AAH05499.1; -;  
 HSSP; P00740; 1EDM.  
 MGD; MGI:1858235; Celar2.  
 InterPro: IPR000152; Asx\_hydroxyl.  
 InterPro: IPR002126; Cadherin.  
 InterPro: IPR000561; EGF-like.  
 InterPro: IPR000742; EGF-2.  
 InterPro: IPR000832; GPCR\_secretin.  
 InterPro: IPR001879; horxn\_receptor.  
 InterPro: IPR002049; Laminin\_EGF.  
 InterPro: IPR001791; Laminin\_G.  
 InterPro: IPR000203; PKD\_cys\_rich.  
 Pfam; PF00002; 7tm\_2; 1.  
 Pfam; PF00028; cadherin; 9.  
 Pfam; PF00008; EGF; 5.  
 Pfam; PF01825; GPS; 1.  
 Pfam; PF02793; HRM; 1.  
 Pfam; PF00054; laminin\_G; 2.  
 PRINTS; PR00205; CADHERIN.  
 PRINTS; PR00011; EGF\_LAMININ.  
 PRINTS; PR00249; GPCR\_SECRETIN.  
 SMART; SM00112; CA; 9.  
 SMART; SM00180; EGF\_Lam; 1.  
 SMART; SM00001; EGF\_like; 6.  
 SMART; SM00303; GPS; 1.  
 SMART; SM00008; HornR; 1.  
 SMART; SM00282; LamG; 2.  
 PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 PROSITE; PS00232; CADHERIN\_1; 6.  
 PROSITE; PS00268; CADHERIN\_2; 9.  
 PROSITE; PS00022; EGF\_1; 6.  
 PROSITE; PS01186; EGF\_2; 2.  
 PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.  
 PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2\_2; FALSE\_NEG.  
 PROSITE; PS00227; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 PROSITE; PS00261; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.  
 PROSITE; PS00221; GPS; 1.  
 PROSITE; PS00025; LAM\_G\_DOMAIN; 2.  
 G-protein coupled receptor; Transmembrane; Glycoprotein; Repeat;  
 EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;  
 Developmental protein; Hydroxylation; Signal.  
 SIGNAL 1 31 POTENTIAL.  
 CHAIN 32 2920 CADHERIN EGF LAG SEVEN-PASS G-TYPE  
 RECEPTOR 2.

FT	DOMAIN	32	2381	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2382	2402	1 (POTENTIAL).
FT	DOMAIN	2403	2414	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2415	2434	2 (POTENTIAL).
FT	DOMAIN	2435	2439	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2440	2460	3 (POTENTIAL).
FT	DOMAIN	2461	2481	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2482	2502	4 (POTENTIAL).
FT	DOMAIN	2503	2519	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2520	2540	5 (POTENTIAL).
FT	DOMAIN	2541	2564	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2565	2585	6 (POTENTIAL).
FT	DOMAIN	2586	2592	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2593	2613	7 (POTENTIAL).
FT	DOMAIN	2614	2920	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	182	289	CADHERIN 1.
FT	DOMAIN	290	399	CADHERIN 2.
FT	DOMAIN	400	506	CADHERIN 3.
FT	DOMAIN	507	611	CADHERIN 4.
FT	DOMAIN	612	713	CADHERIN 5.
FT	DOMAIN	714	816	CADHERIN 6.
FT	DOMAIN	817	922	CADHERIN 7.
FT	DOMAIN	923	1024	CADHERIN 8.
FT	DOMAIN	1029	1147	CADHERIN 9.
FT	DOMAIN	1229	1287	EGF-LIKE 1 (ATYPICAL).
FT	DOMAIN	1289	1319	EGF-LIKE 2, CALCIUM-BINDING.
FT	DOMAIN	1329	1367	EGF-LIKE 3, CALCIUM-BINDING.
FT	DOMAIN	1368	1572	LAMININ G-LIKE 1.
FT	DOMAIN	1575	1611	EGF-LIKE 4, CALCIUM-BINDING.
FT	DOMAIN	1615	1792	LAMININ G-LIKE 2.
FT	DOMAIN	1788	1830	EGF-LIKE 5, CALCIUM-BINDING.
FT	DOMAIN	1831	1868	EGF-LIKE 6, CALCIUM-BINDING.
FT	DOMAIN	1884	1923	EGF-LIKE 7, CALCIUM-BINDING.
FT	DOMAIN	1924	1956	EGF-LIKE 8, CALCIUM-BINDING.
FT	DOMAIN	2317	2369	GPS.
FT	DOMAIN	2744	2749	POLY-GLU.
FT	DISULFID	1293	1304	BY SIMILARITY.
FT	DISULFID	1298	1313	BY SIMILARITY.
FT	DISULFID	1315	1318	BY SIMILARITY.
FT	DISULFID	1333	1344	BY SIMILARITY.
FT	DISULFID	1338	1354	BY SIMILARITY.
FT	DISULFID	1356	1366	BY SIMILARITY.
FT	DISULFID	1379	1590	BY SIMILARITY.
FT	DISULFID	1384	1599	BY SIMILARITY.
FT	DISULFID	1601	1610	BY SIMILARITY.
FT	DISULFID	1792	1803	BY SIMILARITY.
FT	DISULFID	1798	1818	BY SIMILARITY.
FT	DISULFID	1820	1829	BY SIMILARITY.
FT	DISULFID	1833	1844	BY SIMILARITY.
FT	DISULFID	1838	1856	BY SIMILARITY.
FT	DISULFID	1858	1867	BY SIMILARITY.
FT	DISULFID	1888	1900	BY SIMILARITY.
FT	DISULFID	1890	1907	BY SIMILARITY.
FT	DISULFID	1909	1922	BY SIMILARITY.
FT	DISULFID	1925	1937	BY SIMILARITY.
FT	DISULFID	1927	1944	BY SIMILARITY.
FT	DISULFID	1946	1955	BY SIMILARITY.
FT	MOD_RES	1592	1592	HYDROXYLATION (POTENTIAL).
FT	CARBOHYD	486	486	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	558	558	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	702	702	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1037	1037	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1077	1077	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1183	1183	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1213	1213	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1502	1502	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1566	1566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1742	1742	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1828	1828	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1901	1901	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2025	2025	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2044	2044	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2062	2062	N-LINKED (GLCNAC. . .) (POTENTIAL).



```
FT CARBOHYD 2324 2324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2346 2346 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 2199 2199 V -> L (IN REF. 3).
FT CONFLICT 2283 2283 A -> V (IN REF. 3).
FT CONFLICT 2535 2535 S -> R (IN REF. 1).
FT CONFLICT 2571 2571 L -> R (IN REF. 2).
FT CONFLICT 2639 2639 Y -> S (IN REF. 3).

Query Match 85.7%; Score 48; DB 1; Length 2920;
Best Local Similarity 70.0%; Pred. No. 1.7;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSDXC 10
DB 1820 CVLGYDNC 1829
||:||||:|

RESULT 7
NRG2_MOUSE
ID NRG2_MOUSE STANDARD; PRT; 756 AA.
AC P56974;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Pro-neuregulin-2 precursor (Pro-NRG2) [Contains: Neuregulin-2 (NRG-2)
DE (Divergent of neuregulin 1) (DON-1)].
GN NRG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=97311398; PubMed=9168115;
RA Carraway K.L. III, Weber J.L., Unger M.J., Ledesma J., Yu N.,
RA Gassmann M., Lai C.;
RT "Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine
RT kinases."
RL Nature 387:512-516(1997).
RN [2]
RP SEQUENCE OF 150-756 FROM N.A. (ISOFORMS DON-1M AND DON-1S).
RC TISSUE=Choroid plexus;
RX MEDLINE=97342638; PubMed=9199335;
RA Busfield S.J., Michnick D.A., Chickering T.W., Revett T.L., Ma J.,
RA Woolf E.A., Comrack C.A., Dussault B.J., Woolf J., Goodearl A.D.J.,
RA Gearing D.P.;
RT "Characterization of a neuregulin-related gene, Don-1, that is highly
RT expressed in restricted regions of the cerebellum and hippocampus."
RL Mol. Cell. Biol. 17:4007-4014(1997).
CC -|- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
CC RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
CC RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC ACTIVATION OF THE ERBB RECEPTORS. MAY ALSO PROMOTE THE
CC HETERODIMERIZATION WITH THE EGF RECEPTOR.
CC -|- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
CC A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
CC MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC -|- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS: DON-1M, DON-1S/NRG2-5,
CC NRG2-10 AND NRG2-16A (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE
CC SPLICING.
CC -|- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN THE BRAIN, WITH LOWER
CC LEVELS IN THE LUNG. IN THE CEREBELLUM, FOUND IN GRANULE AND
CC PURKINJE CELLS.
CC -|- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC DIMERIZATION (BY SIMILARITY).
CC -|- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC DOMAIN (BY SIMILARITY).
CC -|- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC FORM (BY SIMILARITY).
```

```
CC -|- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
CC SIMILARITY).
CC -|- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
DR HSSP; Q12784; IHRE.
DR MGD; MGI:1098246; Nrg2.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR002154; Neuregulin.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM01181; EGF; 1.
DR SMART; SM00408; IGF2; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW Transmembrane; Multigene family; Alternative splicing.
FT PROPEP 1 19 BY SIMILARITY.
FT CHAIN 20 756 PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.
FT CHAIN 20 314 NEUREGULIN-2.
FT DOMAIN 20 315 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 316 336 INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT DOMAIN 337 756 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 158 226 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 238 248 SER/THR-RICH.
FT DOMAIN 249 290 EGF-LIKE.
FT DOMAIN 627 633 POLY-PRO.
FT DISULFID 165 219 BY SIMILARITY.
FT DISULFID 253 267 BY SIMILARITY.
FT DISULFID 261 278 BY SIMILARITY.
FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 280 280 C -> G (IN ISOFORM NRG2-10).
FT VARSPLIC 281 756 MISSING (IN ISOFORM NRG2-10).
FT VARSPLIC 282 330 CVYTGDRCCQFAMVNFKHLGFEKAEELQKRVLTITGI
FT VARSPLIC 331 756 PGTGVSSQWSTSPSTLDN (IN ISOFORM DON-1S).
FT VARSPLIC 282 307 MISSING (IN ISOFORM DON-1S).
FT VARSPLIC 282 307 VGYTGRCCQFAMVNFKHLGFEKAE -> NGFTGRCLEK
FT VARSPLIC 282 307 LPLRLYMDPKQK (IN ISOFORM DON-1M).
SQ SEQUENCE 756 AA; 82213 MW; 51D85DC918BE678E CRC64;

Query Match 82.1%; Score 46; DB 1; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSDXC 10
DB 280 CPVGYTGDRC 289
|:||||:|

RESULT 8
EGF_RAT
ID EGF_RAT STANDARD; PRT; 1133 AA.
AC P07522; Q63183;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal
DE growth factor].
GN EGF.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
```

RC TISSUE-Kidney;  
RX MEDLINE=92398779; PubMed=1524680;  
RA Price P.M., Saggi S.J., Safirstein R.;  
RT "Cloning and sequencing of the rat preproepidermal growth factor  
RL cDNA: comparison with mouse and human sequences.";  
RN DNA Cell Biol. 11:481-487(1992).  
RP [2]  
RN REVISIONS.  
RC TISSUE-Kidney;  
RA Price P.M.;  
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 974-1021.  
RX MEDLINE=86081810; PubMed=3000782;  
RA Simpson R.J., Smith J.A., Moritz R.L., O'Hare M.J., Rudland P.S.,  
RL Morrison J.R., Lloyd C.J., Grego B., Burgess A.W., Nice E.C.;  
RT "Rat epidermal growth factor: complete amino acid sequence. Homology  
RT with the corresponding murine and human proteins; isolation of a form  
RL truncated at both ends with full in vitro biological activity.";  
RL Eur. J. Biochem. 153:629-637(1985).  
RN [4]  
RP SEQUENCE OF 994-1108 FROM N.A.  
RX STRAIN-Sprague-Dawley; TISSUE-Kidney;  
RA MEDLINE=89016634; PubMed=3282867;  
RL Dorow D.S., Simpson R.J.;  
RT "Cloning and sequence analysis of a cDNA for rat epidermal growth  
RL factor.";  
RL Nucleic Acids Res. 16:9338-9338(1988).  
CC -1- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS  
CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME  
CC FIBROBLASTS IN CELL CULTURE.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U04842; AAB60436.1; -;  
DR EMBL; X12748; CAA31241.1; -;  
DR PIR; A25425; EGRF.  
DR HSSP; P01132; LEGF.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001336; EGF\_1.  
DR InterPro: IPR001881; EGF-Ca.  
DR InterPro: IPR000033; Ldl\_receptor\_rep.  
DR Pfam; PF00008; EGF; 7.  
DR Pfam; PF00058; ldl\_recept\_b; 7.  
DR PRINTS; PR00009; EGF\_TCF.  
DR SMART; SM00179; EGF\_CA; 2.  
DR SMART; SM00001; EGF\_like; 6.  
DR SMART; SM00135; LY; 8.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 3.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 6.  
DR PROSITE; PS01187; EGF\_CA; 3.  
KW EGF-like domain; Repeat; growth factor; Transmembrane; Glycoprotein;  
Signal.  
FT SIGNAL  
FT CHAIN 21  
FT CHAIN 22 1133  
FT CHAIN 974 1026  
FT DOMAIN 22 1035  
FT TRANSMEM 1036 1057  
FT DOMAIN 1058 1133  
FT DOMAIN 322 356  
FT DOMAIN 357 397  
FT DOMAIN 398 438  
FT DOMAIN 436 478  
FT DOMAIN 436 478

FT DOMAIN 743 783  
FT DOMAIN 835 873  
FT DOMAIN 874 915  
FT DOMAIN 916 956  
FT DOMAIN 975 1016  
FT DISULFID 361 372  
FT DISULFID 368 381  
FT DISULFID 383 396  
FT DISULFID 402 413  
FT DISULFID 409 422  
FT DISULFID 424 437  
FT DISULFID 440 452  
FT DISULFID 448 462  
FT DISULFID 464 477  
FT DISULFID 474 758  
FT DISULFID 754 767  
FT DISULFID 769 782  
FT DISULFID 839 850  
FT DISULFID 844 859  
FT DISULFID 861 872  
FT DISULFID 944 955  
FT DISULFID 979 993  
FT DISULFID 987 1004  
FT DISULFID 1006 1015  
FT CARBOHYD 105 105  
FT CARBOHYD 118 118  
FT CARBOHYD 149 149  
FT CARBOHYD 405 405  
FT CARBOHYD 930 930  
FT CARBOHYD 941 941  
FT VARIANT 935 935  
FT CONFLICT 1024 1025  
FT CONFLICT 1108 1108  
FT CONFLICT 1133 AA; 124125 MW; C224A302E9578031 CRC64;  
SQ SEQUENCE 1133 AA; 124125 MW; C224A302E9578031 CRC64;  
Query Match 82.1%; Score 46; DB 1; Length 1133;  
Best Local Similarity 70.0%; Pred. No. 1.6;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CVIGYSGDXC 10  
Db 1006 CVIGYIGERC 1015  
||||| |::|  
DL\_DROME STANDARD; PRT; 833 AA.  
AC P10041; Q9VDY2; Q99108;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Neurogenic locus Delta protein precursor.  
GN DL OR CG3619.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
RC TISSUE-Embryo;  
RA Vaessin H., Bremer K.A., Knust E., Campos-Ortega J.A.;  
RT "The neurogenic gene Delta of Drosophila melanogaster is expressed in  
RT neurogenic territories and encodes a putative transmembrane protein  
RL with EGF-like repeats.";  
RL EMBO J. 6:3431-3440(1987).  
RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN-Oregon-R; TISSUE-Embryo;  
RX MEDLINE=89196890; PubMed=3149249;  
RA Kopczynski C.C., Altan A.K., Fechtel K., Kooh P.J., Muskavitch M.A.T.;  
RT "Delta, a Drosophila neurogenic gene, is transcriptionally complex and  
RT encodes a protein related to blood coagulation factors and epidermal  
RT growth factor of vertebrates.";  
RL Genes Dev. 2:1723-1735(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Ceinaker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Frankkoch C., Baldwin D.,  
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclel J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [4]  
RP SEQUENCE OF 422-621 FROM N.A.  
RC TISSUE-Embryo;  
RX MEDLINE=87218537; PubMed=3107986;  
RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D., Vaessin H.,  
RA Campos-Ortega J.A.;  
RT "EGF homologous sequences encoded in the genome of Drosophila  
RT melanogaster, and their relation to neurogenic genes.";  
RL EMBO J. 6:761-766(1987).  
RN [5]  
RP PATTERN OF TRANSCRIPTION, AND CHARACTERIZATION.  
RX MEDLINE=91209246; PubMed=2128477;  
RA Haenlin M., Kramatschek B., Campos-Ortega J.A.;  
RT "The pattern of transcription of the neurogenic gene Delta of  
RT Drosophila melanogaster.";  
RL Development 110:905-914(1990).  
CC -!- FUNCTION: ESSENTIAL FOR PROPER DIFFERENTIATION OF ECTODERM. DL  
CC IS REQUIRED FOR THE CORRECT SEPARATION OF NEURAL AND EPIDERMAL  
CC CELL LINESAGES.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: DETECTED IN ALL AREAS WITH NEUROGENIC  
CC ABILITIES, FOR EXAMPLE THE NEUROGENIC ECTODERM AND THE PRIMORDIA  
CC OF THE SENSE ORGANS. LATER EXPRESSION IS RESTRICTED TO THOSE CELLS  
CC THAT HAVE ADOPTED A NEURAL FATE.

CC -!- DEVELOPMENTAL STAGE: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.  
CC EXPRESSION IS HIGHEST EARLY IN EMBRYONIC DEVELOPMENT (STAGE 5) AND  
CC REDUCES TO A LOW LEVEL DURING LARVAL STAGES.  
CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO  
CC THE INNER PART OF EMBRIO IS ONE OF THE FIRST STEPS OF CNS  
CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE  
CC NEUROGENIC GENES.  
CC -!- MISCELLANEOUS: NOTCH AND SERRATE MAY INTERACT AT THE PROTEIN  
CC LEVEL, IT IS CONCEIVABLE THAT THE SERRATE AND DELTA PROTEINS MAY  
CC COMPLETE FOR BINDING WITH THE NOTCH PROTEIN.  
CC -!- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 DSL DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X06289; CAA29617.1; -  
CC EMBL; Y00222; CAA68369.1; -  
CC EMBL; A0003725; AAF55657.1; -  
CC PIR; A05140; CAA28786.1; -  
CC PIR; A26637; S00670.  
CC HSSP; P00740; IEDM.  
CC FlyBase; FBgn0000463; DL.  
CC InterPro; IPR000152; Asx\_hydroxyl.  
CC InterPro; IPR001774; DSL.  
CC InterPro; IPR000561; EGF-like.  
CC InterPro; IPR000742; EGF-2.  
CC InterPro; IPR001881; EGF\_Ca.  
CC InterPro; IPR001438; EGF\_II.  
CC InterPro; IPR002049; Laminin\_EGF.  
CC Pfam; PF00008; EGF; 9.  
CC Pfam; PF01414; DSL; 1.  
CC PRINTS; PR00010; EGFBL00D.  
CC PRINTS; PR00011; EGF\_LAMININ.  
CC SMART; SM00051; DSL; 1.  
CC SMART; SM00179; EGF\_Ca; 4.  
CC SMART; SM00001; EGF-like; 5.  
CC PROSITE; PS00010; ASX\_HYDROXYL; 3.  
CC PROSITE; PS00022; EGF\_1; 9.  
CC PROSITE; PS01186; EGF-2; 9.  
CC PROSITE; PS01187; EGF\_Ca; 2.  
KW Developmental protein; Differentiation; Neurogenesis; Repeat;  
KW Transmembrane; EGF-like domain; Glycoprotein; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 833 NEUROGENIC LOCUS DELTA PROTEIN.  
FT DOMAIN 19 594 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 595 617 POTENTIAL.  
FT DOMAIN 618 833 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 164 226 DSL.  
FT DOMAIN 227 258 EGF-LIKE 1.  
FT DOMAIN 256 289 EGF-LIKE 2.  
FT DOMAIN 291 329 EGF-LIKE 3.  
FT DOMAIN 331 372 EGF-LIKE 4.  
FT DOMAIN 374 416 EGF-LIKE 5.  
FT DOMAIN 418 451 EGF-LIKE 6.  
FT DOMAIN 453 489 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 491 527 EGF-LIKE 8.  
FT DOMAIN 529 565 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).  
FT DISULFID 231 240 BY SIMILARITY.  
FT DISULFID 235 246 BY SIMILARITY.  
FT DISULFID 248 257 BY SIMILARITY.  
FT DISULFID 260 271 BY SIMILARITY.  
FT DISULFID 266 277 BY SIMILARITY.  
FT DISULFID 279 288 BY SIMILARITY.  
FT DISULFID 295 307 BY SIMILARITY.  
FT DISULFID 301 317 BY SIMILARITY.  
FT DISULFID 319 328 BY SIMILARITY.

```

FT DISULFID 335 BY SIMILARITY. 348
FT DISULFID 342 BY SIMILARITY. 360
FT DISULFID 362 BY SIMILARITY. 371
FT DISULFID 378 BY SIMILARITY. 388
FT DISULFID 383 BY SIMILARITY. 404
FT DISULFID 406 BY SIMILARITY. 415
FT DISULFID 422 BY SIMILARITY. 433
FT DISULFID 437 BY SIMILARITY. 439
FT DISULFID 441 BY SIMILARITY. 450
FT DISULFID 457 BY SIMILARITY. 468
FT DISULFID 462 BY SIMILARITY. 477
FT DISULFID 479 BY SIMILARITY. 488
FT DISULFID 495 BY SIMILARITY. 506
FT DISULFID 500 BY SIMILARITY. 515
FT DISULFID 517 BY SIMILARITY. 526
FT DISULFID 533 BY SIMILARITY. 544
FT DISULFID 538 BY SIMILARITY. 553
FT DISULFID 555 BY SIMILARITY. 564
FT CARBOHYD 98 N-LINKED (GLCNAC. . .) (POTENTIAL). 98
FT CARBOHYD 137 N-LINKED (GLCNAC. . .) (POTENTIAL). 137
FT CARBOHYD 167 N-LINKED (GLCNAC. . .) (POTENTIAL). 167
FT CONFLICT 5 K -> N (IN REF. 1). 5
FT CONFLICT 67 V -> L (IN REF. 1). 67
FT CONFLICT 363 A -> R (IN REF. 1). 363
FT CONFLICT 437 GK -> ET (IN REF. 3). 437
FT CONFLICT 443 A -> S (IN REF. 3). 443
FT CONFLICT 459 G -> A (IN REF. 3). 459
FT CONFLICT 490 S -> T (IN REF. 3). 490
FT CONFLICT 591 T -> A (IN REF. 1 AND 3). 591
FT CONFLICT 631 D -> N (IN REF. 1). 631
FT CONFLICT 652 G -> A (IN REF. 1). 652
FT CONFLICT 662 L -> M (IN REF. 1). 662

Query Match 80.4%; Score 45; DB 1; Length 833;
Best Local Similarity 70.0%; Pred. No. 1.8;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGSGDGC 10
DB 319 CAPGYSGDGC 328

RESULT 10
PCGB_RAT STANDARD; PRT; 883 AA.
AC P53068; Q63040; Q62860; Q63513;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Brevican core protein precursor (Brain enriched hyaluronan binding protein) (BEHAB protein).
DE BCAN.
GN Rattus norvegicus (Rat).
OS
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=96070828; PubMed=752978;
RA Seidenbecher C.I., Richter K., Rauch U., Faessler R., Garner C.C.,
RA Gundelfinger E.D.;
FT "Brevican, a chondroitin sulfate proteoglycan of rat brain, occurs as
FT secreted and cell surface glycosylphosphatidylinositol-anchored
FT isoforms.";
RL J. Biol. Chem. 270:27206-27212(1995).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 396-407.
RC TISSUE=Brain;
RX MEDLINE=96074575; PubMed=7488217;
RA Yamada H., Watanabe K., Shimonaka M., Yamasaki M., Yamauchi Y.;
RT "cDNA cloning and the identification of an aggrecanase-like cleavage
RT site in rat brevican.";

```

```

RL Biochem. Biophys. Res. Commun. 216:957-963(1995).
RN [3]
RP SEQUENCE OF 1-423 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=94216386; PubMed=7512973;
RA Jaworski D.M., Kelly G.M., Hockfield S.;
RT "BEHAB, a new member of the proteoglycan tandem repeat family of
RT hyaluronan-binding proteins that is restricted to the brain.";
RL J. Cell Biol. 125:455-509(1994).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE TERMINALLY DIFFERENTIATING AND
CC THE ADULT NERVOUS SYSTEM DURING POSTNATAL DEVELOPMENT. COULD
CC STABILIZE INTERACTIONS BETWEEN HA AND BRAIN PROTEOGLYCAN. THE
CC GPI-ANCHORED ISOFORM MAY FUNCTION AS A CHONDROITIN SULFATE-
CC BEARING CELL SURFACE RECEPTOR.
CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX AND ONE FORM
CC ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SECRETED FORM (SHOWN HERE) AND
CC A GPI-ANCHORED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- DEVELOPMENTAL STAGE: SOLUBLE FORM INCREASES FROM DAY P4 TO P64.
CC GPI-ANCHORED ISOFORM INCREASES AFTER DAY P8.
CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 364
CC ONWARD AND IS SMALLER (371 AA) DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X79881; CAA56255.1; -
DR EMBL; X86406; CAA60160.1; -
DR EMBL; U37142; AAM87847.1; -
DR EMBL; Z28366; CAA82215.1; ALT_FRAME.
DR HSSP; P20693; 1HLJ.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; SUSHI_SCR_CCP.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sush; 1.
DR Pfam; PF00193; Xlink; 2.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00406; IGV; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Lectin; Signal; SUSHI;
KW EGF-like domain; Repeat; Immunoglobulin domain; Alternative splicing;
KW GPI-anchor.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 883 BREVICAN CORE PROTEIN.

```

```

FT DOMAIN 32 157 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 173 250 LINK 1.
FT DOMAIN 271 352 LINK 2.
FT DOMAIN 622 658 EGF-LIKE.
FT DOMAIN 658 786 C-TYPE LECTIN.
FT DOMAIN 787 851 SUSH1.
FT DISULFID 56 136 BY SIMILARITY.
FT DISULFID 178 249 BY SIMILARITY.
FT DISULFID 202 223 BY SIMILARITY.
FT DISULFID 276 351 BY SIMILARITY.
FT DISULFID 300 321 BY SIMILARITY.
FT DISULFID 636 637 BY SIMILARITY.
FT DISULFID 631 646 BY SIMILARITY.
FT DISULFID 648 657 BY SIMILARITY.
FT DISULFID 791 834 BY SIMILARITY.
FT DISULFID 820 847 BY SIMILARITY.
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 625 645 DCIPSPCHNGGTCLEKEGR -> NSAEGSPAFLLFLLL
FT VARSPLIC 646 883 MISSING (IN GPI-ANCHORED ISOFORM).
FT CONFLICT 51 52 AL -> WV (IN REF. 3).
FT CONFLICT 503 503 V -> L (IN REF. 2).
FT CONFLICT 518 519 TV -> PA (IN REF. 2).
FT CONFLICT 526 526 G -> R (IN REF. 2).
FT CONFLICT 541 541 G -> A (IN REF. 2).
FT CONFLICT 556 556 R -> S (IN REF. 2).
FT CONFLICT 573 573 E -> A (IN REF. 2).
FT CONFLICT 583 583 V -> L (IN REF. 2).
FT CONFLICT 649 649 V -> L (IN REF. 2).
FT CONFLICT 670 670 P -> A (IN REF. 2).
FT CONFLICT 738 738 P -> A (IN REF. 2).
FT CONFLICT 809 809 R -> A (IN REF. 2).
SQ SEQUENCE 883 AA; 96057 MW; AC7ACC40CB53ED37 CRC64;

Query Match 80.4%; Score 45; DB 1; Length 883;
Best Local Similarity 70.0%; Pred. No. 1.9;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVVIGSGDXC 10
DB 648 CVPVGGDLG 657

RESULT 11
EGF_HUMAN STANDARD; PRT; 1207 AA.
AC P01133;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal growth factor (Urogastrone)].
GN EGF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=87066721; PubMed=3491360;
RA Bell G.I., Fong N.M., Stempflen M.M., Wormsted M.A., Caput D.,
RA Ku L., Urdea M.S., Rall L.B., Sanchez-Pescador R.;
RT "Human epidermal growth factor precursor: cDNA sequence, expression in vitro and gene organization.";
RL Nucleic Acids Res. 14:8427-8446(1986).
[2]
RN SEQUENCE OF 971-1023.
RX MEDLINE=77117897; PubMed=300079;
RA Gregory H., Preston B.M.;
RT "The primary structure of human urogastrone.";
RL Int. J. Pept. Protein Res. 9:107-118(1977).
```

```

RN [3]
RX SEQUENCE OF 971-1023.
RX MEDLINE=89391964; PubMed=2789514;
RA Furuya M., Akashi S., Hirayama K.;
RT "The primary structure of human EGF produced by genetic engineering, studied by high-performance tandem mass spectrometry.";
RL Biochem. Biophys. Res. Commun. 163:1100-1106(1989).
[4]
RN STRUCTURE BY NMR OF EGF.
RX MEDLINE=92395667; PubMed=1522591;
RA Hommel U., Harvey T.S., Driscoll P.C., Campbell I.D.;
RT "Human epidermal growth factor. High resolution solution structure and comparison with human transforming growth factor alpha.";
RL J. Mol. Biol. 227:271-282(1992).
CC -1- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME FIBROBLASTS IN CELL CULTURE.
CC FIBROBLAST LOCATION: Type I membrane protein.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/ or send an email to license@isb-sib.ch).
CC
EMBL: X04571; CAA28240.1;
PIR: A25331; EGHU.
DR HSSP; P01132; LEGF.
DR Genew; HGNC:3229; EGF.
DR MIM; 131530;
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001336; EGF_1.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000033; Ldl_receptor_rep.
DR Pfam; PF00008; EGF; 9.
DR Pfam; PF00058; ldl_recept_b; 7.
DR PRINTS; PR00009; EGFTGF.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF_like; 7.
DR SMART; SM00135; LY; 8.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 3.
KW EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein; Signal; Polymorphism.
FT SIGNAL 1 22
FT CHAIN 23 1207 PRO-EPIDERMAL GROWTH FACTOR.
FT CHAIN 971 1023 EPIDERMAL GROWTH FACTOR.
FT DOMAIN 23 1032 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1033 1053 POTENTIAL.
FT DOMAIN 1054 1207 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 314 355 EGF-LIKE 1.
FT DOMAIN 356 396 EGF-LIKE 2.
FT DOMAIN 397 437 EGF-LIKE 3.
FT DOMAIN 435 477 EGF-LIKE 4.
FT DOMAIN 741 781 EGF-LIKE 5.
FT DOMAIN 831 869 EGF-LIKE 6.
FT DOMAIN 912 952 EGF-LIKE 7.
FT DOMAIN 972 1013 EGF-LIKE 8.
FT DISULFID 318 330 BY SIMILARITY.
FT DISULFID 325 339 BY SIMILARITY.
FT DISULFID 341 354 BY SIMILARITY.
FT DISULFID 360 371 BY SIMILARITY.
FT DISULFID 367 380 BY SIMILARITY.
FT DISULFID 382 395 BY SIMILARITY.
FT DISULFID 401 412 BY SIMILARITY.
FT DISULFID 408 421 BY SIMILARITY.
```





"Proteoglycan expression in the normal rat kidney."  
[4]  
Sequence of 2535-2738 FROM N.A.  
RC STRAIN-Sprague-Dawley; TISSUE=Lung;  
RA Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;  
RT "Molecular cloning and characterization of two developmentally  
regulated genes in rat lung."  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases  
CC -!- FUNCTION: May play a role in intercellular signaling and in  
connecting cells with the extracellular matrix. May take part in  
the regulation of cell motility, growth and differentiation. Binds  
hyaluronic acid.  
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
CC -!- ALTERNATIVE PRODUCTS: At least 3 isoforms; V0 (shown here), V3 and  
Vint; are produced by alternative splicing.  
CC -!- TISSUE SPECIFICITY: In kidney is expressed in the papillary area,  
but not in glomeruli.  
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development  
(By similarity).  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.  
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL; AF062402; AAC40166.1; -  
DR EMBL; U75306; AAB51125.1; -  
DR EMBL; AF084544; AAD48544.1; -  
DR EMBL; AF072892; AAC26116.1; -  
DR EMBL; AY007691; AAG16631.1; -  
DR HSP; P01132; 1EPG.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR000742; EGF-2.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR003599; Ig.  
DR InterPro: IPR003600; Ig\_Like.  
DR InterPro: IPR001304; Lectin\_C.  
DR InterPro: IPR000538; Link.  
DR PRINTS; PR01265; LINKMODULE.  
DR SMART; SM00032; CCP; 1.  
DR SMART; SM00034; CLECT; 1.  
DR SMART; SM00181; EGF; 2.  
DR SMART; SM00179; EGF\_CA; 2.  
DR SMART; SM00409; IG; 1.  
DR SMART; SM00410; IG\_Like; 1.  
DR SMART; SM00445; LINK; 2.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
DR PROSITE; PS00041; C-TYPE LECTIN\_2; 1.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS00186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 1.  
DR PROSITE; PS01241; LINK; 2.  
DR Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;  
Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
Hyaluronic acid; Alternative splicing.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 2738 VERSICAN CORE PROTEIN.  
FT NON\_CONS 348 349  
FT DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 167 244 LINK 1.

FT	DOMAIN	265	346	LINK 2.
FT	DOMAIN	<349	695	GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
FT	DOMAIN	696	2431	GAG-BETA.
FT	DOMAIN	2431	2467	EGF-LIKE 1.
FT	DOMAIN	2469	2505	EGF-LIKE 2, CALCIUM-BINDING.
FT	DOMAIN	2632	2632	C-TYPE LECTIN.
FT	DOMAIN	2637	2695	SUSHI.
FT	DISULFID	44	130	BY SIMILARITY.
FT	DISULFID	172	243	BY SIMILARITY.
FT	DISULFID	196	217	BY SIMILARITY.
FT	DISULFID	270	345	BY SIMILARITY.
FT	DISULFID	294	315	BY SIMILARITY.
FT	DISULFID	2435	2446	BY SIMILARITY.
FT	DISULFID	2440	2455	BY SIMILARITY.
FT	DISULFID	2457	2466	BY SIMILARITY.
FT	DISULFID	2473	2484	BY SIMILARITY.
FT	DISULFID	2478	2493	BY SIMILARITY.
FT	DISULFID	2495	2504	BY SIMILARITY.
FT	DISULFID	2511	2522	BY SIMILARITY.
FT	DISULFID	2539	2631	BY SIMILARITY.
FT	DISULFID	2607	2623	BY SIMILARITY.
FT	DISULFID	2638	2681	BY SIMILARITY.
FT	DISULFID	2667	2694	BY SIMILARITY.
FT	CARBOHYD	57	57	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	330	330	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	692	692	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	758	758	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	802	802	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	805	805	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1257	1257	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1435	1435	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1633	1633	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1660	1660	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1684	1684	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1738	1738	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1848	1848	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2004	2004	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2409	2409	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2711	2711	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2721	2721	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	349	2431	MISSING (IN ISOFORM V3).
FT	VARSPLIC	2697	2738	PSAYQRTYKSKYKLNSSVKNDSINTSKHHRWSRWQETR R -> RKWSFKKQPCFNKY (IN ISOFORM VINT).
FT	CONFLICT	2535	2539	AEEEC -> NSARG (IN REF 4).
FT	SEQUENCE	2738	AA; 300004	AA; 12CA626D58DBD8C6A CRC64;
FT	SEQUENCE	2738	AA; 300004	AA; 12CA626D58DBD8C6A CRC64;

Query Match 80.4%; Score 45; DB 1; Length 2738;  
Best Local Similarity 70.0%; Pred. No. 5.6;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY	1	CVIGYSGDXC	10
Db	2457	CAPGYSGDQC	2466

RESULT 14  
PGCV\_MOUSE  
ID PGCV\_MOUSE STANDARD; PRT; 3358 AA.  
AC Q62059; Q62058; O9CUU0;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Versican core protein precursor (Large fibroblast proteoglycan)  
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M).  
GN CSPG2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS V0; V1 AND V2).  
RC STRAIN=C57BL/6, and Swiss Webster; TISSUE=Brain;





FT VARSPLIC 348 348 P -> R (IN ISOFORM V1 AND ISOFORM V3).  
 FT VARSPLIC 349 1308 MISSING (IN ISOFORM V1).  
 FT VARSPLIC 1309 3052 MISSING (IN ISOFORM V2).  
 FT VARSPLIC 349 3052 MISSING (IN ISOFORM V3).  
 FT CONFLICT 126 126 A -> G (IN REF. 3).  
 FT CONFLICT 348 348 MISSING (IN REF. 3).  
 FT CONFLICT 1658 1658 I -> T (IN REF. 3).  
 FT CONFLICT 1674 1680 TVNSNS -> OFGIOTA (IN REF. 3).  
 SQ SEQUENCE 3358 AA; 36693 MW; 071B80026BC0762D CRC64;

Query Match 80.4%; Score 45; DB 1; Length 3358;  
 Best Local Similarity 70.08; Pred. No. 6.8;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVIGYSGDXC 10  
 Db 3078 CAPGYSGDQC 3087  
 I |||||

RESULT 15  
 FBPI\_STRPU STANDARD; PRT: 1064 AA.  
 AC P10079;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Fibropellin I precursor (Epidermal growth factor-related protein 1)  
 DE (UEGF-1).  
 GN EGF1.  
 OS Strongylocentrotus purpuratus (Purple sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;  
 OC Strongylocentrotus.  
 OX NCBI\_TaxID=7668;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90112459; PubMed=2514273;  
 RA Delgadillo-Reynoso M.G., Rollo D.R., Hursh D.A., Raff R.A.;  
 RT "Structural analysis of the UEGF gene in the sea urchin  
 strongylocentrotus purpuratus reveals more similarity to vertebrate  
 than to invertebrate genes with EGF-like repeats.";  
 RL J. Mol. Evol. 29:314-327(1989).  
 RN [2]  
 RP SEQUENCE OF 2779-476 AND 781-1064 FROM N.A.  
 RX MEDLINE=87319677; PubMed=3498216;  
 RA Hursh D.A., Andrews M.E., Raff R.A.;  
 RT "A sea urchin gene encodes a polypeptide homologous to epidermal  
 growth factor.";  
 RL Science 237:1487-1490(1987).  
 RN [3]  
 RP AVIDIN-LIKE DOMAIN.  
 RX MEDLINE=89196806; PubMed=2784773;  
 RA Hunt L.T., Barker W.C.;  
 RT "Avidin-like domain in an epidermal growth factor homolog from a sea  
 urchin.";  
 RL FASEB J. 3:1760-1764(1989).  
 RN [4]  
 RP CHARACTERIZATION.  
 RX MEDLINE=91285254; PubMed=2060714;  
 RA Bisgrove B.W., Andrews M.E., Raff R.A.;  
 RT "Fibropellins, products of an EGF repeat-containing gene, form a  
 unique extracellular matrix structure that surrounds the sea urchin  
 embryo.";  
 RL Dev. Biol. 146:89-99(1991).  
 RL Dev. Biol. 146:89-99(1991).  
 CC -1- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR  
 MATRIX.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM  
 OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER  
 THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE  
 EMBRYOS AND EARLY LARVAE.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1A (SHOWN HERE) AND 1B; ARE  
 PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (1B) LACKS 8 EGF  
 REPEATS.

CC -1- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND  
 CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN  
 CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS  
 CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY  
 CC AND ZYGOTICALLY.  
 CC -1- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
 CC -1- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR  
 CC TO AVIDIN/STREPTAVIDIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L08692; AAA62164.1; -;  
 DR EMBL: L08692; AAA62163.1; -;  
 DR EMBL: X17530; CAA35571.1; -;  
 DR EMBL: M17421; AAA30050.1; -;  
 DR EMBL: X17533; CAA35573.1; -;  
 DR PIR: A29316; A29316.  
 DR HSP: P01132; IEGF.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000088; Avidin.  
 DR InterPro: IPR000859; CUB\_domain.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF 2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001438; EGF\_II.  
 DR InterPro: IPR002049; Laminin\_EGF.  
 DR Pfam: PF00008; EGF; 21.  
 DR Pfam: PF00431; CUB; 1.  
 DR Pfam: PF01382; Avidin; 1.  
 DR PRINTS: PR00010; EGFBLD.  
 DR PRINTS: PR00011; EGFLAMININ.  
 DR SMART: SM00042; CUB; 1.  
 DR SMART: SM00179; EGF\_Ca; 18.  
 DR SMART: SM00001; EGF\_like; 3.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 19.  
 DR PROSITE: PS00022; EGF\_1; 19.  
 DR PROSITE: PS00577; AVIDIN; 1.  
 DR PROSITE: PS01180; CUB; 1.  
 DR PROSITE: PS01185; EGF\_2; 19.  
 DR PROSITE: PS01187; EGF\_Ca; 18.  
 KW Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;  
 KW Glycoprotein; Calcium-binding.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 1064 FIBROPELLIN I.  
 FT DOMAIN 20 55 EGF-LIKE 1.  
 FT DOMAIN 62 175 CUB.  
 FT DOMAIN 176 212 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 366 402 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 442 478 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 480 516 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 518 554 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 556 592 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 594 630 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 632 668 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 670 706 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 708 744 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 746 782 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 784 820 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 822 858 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 860 896 EGF-LIKE 20.  
 FT DOMAIN 898 934 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).

FT	DOMAIN	936	1064	AVIDIN-LIKE.
FT	DISULFID	23	34	BY SIMILARITY.
FT	DISULFID	28	43	BY SIMILARITY.
FT	DISULFID	45	54	BY SIMILARITY.
FT	DISULFID	180	191	BY SIMILARITY.
FT	DISULFID	185	200	BY SIMILARITY.
FT	DISULFID	202	211	BY SIMILARITY.
FT	DISULFID	218	229	BY SIMILARITY.
FT	DISULFID	223	238	BY SIMILARITY.
FT	DISULFID	240	249	BY SIMILARITY.
FT	DISULFID	256	267	BY SIMILARITY.
FT	DISULFID	261	276	BY SIMILARITY.
FT	DISULFID	278	287	BY SIMILARITY.
FT	DISULFID	294	305	BY SIMILARITY.
FT	DISULFID	299	314	BY SIMILARITY.
FT	DISULFID	316	325	BY SIMILARITY.
FT	DISULFID	332	343	BY SIMILARITY.
FT	DISULFID	337	352	BY SIMILARITY.
FT	DISULFID	354	363	BY SIMILARITY.
FT	DISULFID	370	381	BY SIMILARITY.
FT	DISULFID	375	390	BY SIMILARITY.
FT	DISULFID	392	401	BY SIMILARITY.
FT	DISULFID	408	419	BY SIMILARITY.
FT	DISULFID	413	428	BY SIMILARITY.
FT	DISULFID	430	439	BY SIMILARITY.
FT	DISULFID	446	457	BY SIMILARITY.
FT	DISULFID	451	466	BY SIMILARITY.
FT	DISULFID	468	477	BY SIMILARITY.
FT	DISULFID	484	495	BY SIMILARITY.
FT	DISULFID	489	504	BY SIMILARITY.
FT	DISULFID	506	515	BY SIMILARITY.
FT	DISULFID	522	533	BY SIMILARITY.
FT	DISULFID	527	542	BY SIMILARITY.
FT	DISULFID	544	553	BY SIMILARITY.
FT	DISULFID	560	571	BY SIMILARITY.
FT	DISULFID	565	580	BY SIMILARITY.
FT	DISULFID	582	591	BY SIMILARITY.
FT	DISULFID	598	609	BY SIMILARITY.
FT	DISULFID	603	618	BY SIMILARITY.
FT	DISULFID	620	629	BY SIMILARITY.
FT	DISULFID	636	647	BY SIMILARITY.
FT	DISULFID	641	656	BY SIMILARITY.
FT	DISULFID	658	667	BY SIMILARITY.
FT	DISULFID	674	685	BY SIMILARITY.
FT	DISULFID	679	694	BY SIMILARITY.
FT	DISULFID	696	705	BY SIMILARITY.
FT	DISULFID	712	723	BY SIMILARITY.
FT	DISULFID	717	732	BY SIMILARITY.
FT	DISULFID	734	743	BY SIMILARITY.
FT	DISULFID	750	761	BY SIMILARITY.
FT	DISULFID	755	770	BY SIMILARITY.
FT	DISULFID	772	781	BY SIMILARITY.
FT	DISULFID	788	799	BY SIMILARITY.
FT	DISULFID	793	808	BY SIMILARITY.
FT	DISULFID	810	819	BY SIMILARITY.
FT	DISULFID	826	837	BY SIMILARITY.
FT	DISULFID	831	846	BY SIMILARITY.
FT	DISULFID	848	857	BY SIMILARITY.
FT	DISULFID	864	875	BY SIMILARITY.
FT	DISULFID	869	884	BY SIMILARITY.
FT	DISULFID	886	895	BY SIMILARITY.
FT	DISULFID	902	913	BY SIMILARITY.
FT	DISULFID	907	922	BY SIMILARITY.
FT	DISULFID	924	933	BY SIMILARITY.
FT	CARBOHYD	30	30	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	136	136	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	851	851	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	477	780	MISSING (IN ISOFORM IB).
FT	CONFLICT	279	279	L -> S (IN REF. 2)
SQ	SEQUENCE	1064	AB; 112072 MW; 2E569CA012ED6D09 CRC64;	

Query Match 78.6%; Score 44; DB 1; Length 1064;  
Best Local Similarity 60.0%; Pred. No. 3.4;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Oy 1 CVIGYSGDXC 10  
| :|:|:|:|  
Db 316 CPLGFGSDNC 325  
Search completed: July 2, 2003, 07:31:33  
Job time : 11.3333 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 41.3333 Seconds  
(without alignments)  
49.850 Million cell updates/sec

Title: US-09-673-785D-9  
Perfect score: 56  
Sequence: 1 CVIGYSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp-archaea:\*  
2: sp-bacteria:\*  
3: sp-fungi:\*  
4: sp-human:\*  
5: sp-invertebrate:\*  
6: sp-mammal:\*  
7: sp-mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-virus:\*  
13: sp-vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	1200	11	Q8VD07 mus musculus
2	51	91.1	2146	5	Q9VC97 drosophila
3	51	91.1	4006	11	O35452 mus musculus
4	51	91.1	4114	11	O54796 mus musculus
5	49	87.5	372	5	Q21756 caenorhabdi
6	48	85.7	191	5	Q8T521 caenorhabdi
7	48	85.7	191	5	Q8T520 caenorhabdi
8	48	85.7	193	5	Q8T522 caenorhabdi
9	48	85.7	193	5	Q8STG0 caenorhabdi
10	48	85.7	194	5	Q8T523 caenorhabdi
11	48	85.7	963	5	Q9GPM9 caenorhabdi
12	48	85.7	1162	5	Q8WTP0 halocynthia
13	48	85.7	1270	5	Q9GPN0 caenorhabdi
14	48	85.7	1531	11	Q9WVB5 mus musculus
15	48	85.7	2920	11	Q9ROM0 mus musculus
16	46	82.1	1530	11	Q9WUG5 rattus norv

17	46	82.1	1531	11	O88279
18	45	80.4	89	6	Q28867
19	45	80.4	403	5	O18375
20	45	80.4	559	5	Q9VZ44
21	45	80.4	775	5	Q8T314
22	45	80.4	1722	5	Q19350
23	44	78.6	165	13	Q90XX0
24	44	78.6	264	5	Q20043
25	44	78.6	2352	5	O61240
26	44	78.6	4135	6	O18977
27	44	78.6	4288	4	Q9NPK9
28	43	76.8	225	5	Q8T2M9
29	43	76.8	399	5	Q20459
30	43	76.8	464	5	Q95RM9
31	43	76.8	615	13	O57409
32	43	76.8	664	4	Q9U1L7
33	43	76.8	752	13	O42374
34	43	76.8	969	4	Q96KG6
35	43	76.8	1515	13	Q9DE37
36	43	76.8	1534	4	O75093
37	43	76.8	2656	5	Q9GNU3
38	43	76.8	3704	5	P91904
39	42	75.0	106	4	Q9UN93
40	42	75.0	308	4	Q9UN94
41	42	75.0	330	4	Q9NT67
42	42	75.0	366	4	Q9UN95
43	42	75.0	522	4	O8TEP7
44	42	75.0	548	5	O9XUC0
45	42	75.0	790	5	Q9GSG3

## ALIGNMENTS

### RESULT 1

Q8VD07  
ID Q8VD07 PRELIMINARY; PRT; 1200 AA.  
AC Q8VD07;  
DT 01-MAR-2002 (TRENBLrel. 20, Created)  
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE Similar to epidermal growth factor.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RL Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC017681; AAH17681.1; -;  
DR InterPro; IPR000152; ASX\_hydroxyl.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001336; EGF 1.  
DR InterPro; IPR001881; EGF Ca.  
DR InterPro; IPR000033; Ldl\_receptor\_rep.  
DR Pfam; PF00008; EGF; 8.  
DR Pfam; PF00058; ldl\_recept\_b; 7.  
DR PRINTS; PR00009; EGF.TGF.  
DR SMART; SM00181; EGF; 9.  
DR SMART; SM00179; EGF CA; 8.  
DR SMART; SM00135; LY; 9.  
DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_3.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
DR PROSITE; PS01186; EGF\_2; UNKNOWN\_6.  
DR PROSITE; PS01187; EGF CA; UNKNOWN\_3.  
SQ SEQUENCE 1200 AA; 131317 MW; B6A44F7294746476 CRC64;

Query Match 100.0%; Score 56; DB 11; Length 1200;  
Best Local Similarity 90.0%; Pred. No. 0.12;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CVIGSGDXC 10
Db 992 CVIGSGDRC 1001

RESULT 2
Q9VC97 09VC97 PRELIMINARY; PRT; 2146 AA.
AC Q9VC97;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CRB protein.
GN CRB OR CG6383.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J., Blake J., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liao X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR HSSP; P00740; 1EDM.
DR FlyBase; FBgn000368; crb.
DR InterPro; IPR000152; Asx_Hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF; 26.
DR Pfam; PF00054; Laminin_G; 3.

Query Match 91.1%; Score 51; DB 11; Length 4006;
Best Local Similarity 70.0%; Pred. No. 3.4;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10
Db 532 CAVIGSGDRC 541

RESULT 4
Q54796 054796 PRELIMINARY; PRT; 4114 AA.
AC Q54796;
DT 01-JUN-1998 (TREMBlrel. 06, Created)

```

```
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
EN Tenascin-X.
GN TNX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F1 BETWEEN C57BL6, AND CBA;
RX MEDLINE=99014231; PubMed=9795100;
RA Ikuta T., Sogawa N., Ariga Y., Ikemura T., Matsumoto K.;
RT "Structural analysis of mouse tenascin-X: evolutionary aspects of
RT reduplication of FNIII repeats in the tenascin gene family.";
RL Gene 217:1-13(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=F1 BETWEEN C57BL6, AND CBA;
RX MEDLINE=94216385; PubMed=7512972;
RA Matsumoto K., Saga Y., Ikemura T., Sakakura T., Chiquet-Ehrismann R.;
RT "The Distribution of tenascin-X is distinct and often reciprocal to
RT that of tenascin-C.";
RL J. Cell Biol. 125:483-493(1994).
DR EMBL; AB010266; BAA24436.1; -.
DR HSSP; P02671; 1FZD.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00008; EGF; 15.
DR Pfam; PF00147; fibrinogen_C; 1.
DR Pfam; PF00041; fn3; 31.
DR SMART; SM00181; EGF; 7.
DR SMART; SM00001; EGF-like; 2.
DR SMART; SM00186; FBG; 1.
DR SMART; SM00060; FN3; 24.
DR PROSITE; PS00022; EGF_1; UNKNOWN_19.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 4114 AA; 447268 MW; 45FE7AD5145981A1 CRC64;

Query Match 91.1%; Score 51; DB 11; Length 4114;
Best Local Similarity 70.0%; Pred. No. 3.5;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
DB 532 CAVGYSGDPC 541

RESULT 5
ID Q21756 PRELIMINARY; PRT; 372 AA.
AC Q21756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 39.1 kDa protein.
GN R0566.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. the C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Murray J., Le T.T.;
RT "The sequence of C. elegans cosmid R05G6.";
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U58746; AAB00626.1; -.
DR HSSP; P01132; 1A3P.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 5.
DR SMART; SM00181; EGF; 7.
DR PROSITE; PS00022; EGF_1; UNKNOWN_7.
DR PROSITE; PS01186; EGF_2; 6.
DR EGF-like domain; Glycoprotein; Hypothetical protein.
KW EGF-like domain; Glycoprotein; Hypothetical protein.
SQ SEQUENCE 372 AA; 39085 MW; DB36AB251EEB6884 CRC64;

Query Match 87.5%; Score 49; DB 5; Length 372;
Best Local Similarity 70.0%; Pred. No. 0.67;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
DB 233 CYLGSGDKC 242

RESULT 6
ID Q8T521 PRELIMINARY; PRT; 191 AA.
AC Q8T521;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE G1p-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PB800;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491462; AAM09702.1; -.
FT NON_TER 1 191
FT NON_TER 1 191
SQ SEQUENCE 191 AA; 20977 MW; 2EBB21B5FFA46470 CRC64;

Query Match 85.7%; Score 48; DB 5; Length 191;
Best Local Similarity 70.0%; Pred. No. 0.52;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
DB 53 CPLGYSGDYC 62

RESULT 7
ID Q8T520 PRELIMINARY; PRT; 191 AA.
AC Q8T520;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE G1p-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
```

OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6238;  
 RN [1]  
 RC SEQUENCE FROM N.A.

RP STRAIN=VT847;  
 RT Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;  
 "Levels of DNA polymorphism vary with mating system in the nematode  
 genus Caenorhabditis.";  
 RL Genetics 0:0-0(2002).  
 DR EMBL; AF491463; AAM09703.1; -.  
 FT NON\_TER 1 191  
 SQ SEQUENCE 191 AA; 20943 MW; 2EEB21B5F47FB470 CRC64;

Query Match 85.7%; Score 48; DB 5; Length 191;  
 Best Local Similarity 70.0%; Pred. No. 0.52;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 CVIGYSGDXC 10  
 Db 53 CPLGYSGDYC 62  
 I :|||||:I

## RESULT 8

O8T522 PRELIMINARY; PRT; 193 AA.  
 ID O8T522  
 AC O8T522;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DE Glp-1 (Fragment).  
 OS Caenorhabditis briggsae.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6238;  
 RN [1]  
 RC SEQUENCE FROM N.A.

RP STRAIN=AF16;  
 RT Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;  
 "Levels of DNA polymorphism vary with mating system in the nematode  
 genus Caenorhabditis.";  
 RL Genetics 0:0-0(2002).  
 DR EMBL; AF491460; AAM09700.1; -.  
 FT NON\_TER 1 193  
 SQ SEQUENCE 193 AA; 21178 MW; 42135BB8B8BE02C5 CRC64;

Query Match 85.7%; Score 48; DB 5; Length 193;  
 Best Local Similarity 70.0%; Pred. No. 0.52;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 CVIGYSGDXC 10  
 Db 55 CPLGYSGDYC 64  
 I :|||||:I

## RESULT 9

O8STG0 PRELIMINARY; PRT; 193 AA.  
 ID O8STG0;  
 AC O8STG0;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DE Glp-1 (Fragment).  
 OS Caenorhabditis briggsae.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6238;  
 RN [1]  
 RC SEQUENCE FROM N.A.

RP STRAIN=HK104, AND HK105;  
 RT Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;  
 "Levels of DNA polymorphism vary with mating system in the nematode  
 genus Caenorhabditis.";  
 RL Genetics 0:0-0(2002).  
 DR EMBL; AF491463; AAM09703.1; -.  
 FT NON\_TER 1 191  
 SQ SEQUENCE 191 AA; 20943 MW; 2EEB21B5F47FB470 CRC64;

RT genus Caenorhabditis.";  
 RL Genetics 0:0-0(2002).  
 DR EMBL; AF491459; AAM09699.1; -.  
 DR EMBL; AF491461; AAM09701.1; -.  
 FT NON\_TER 1 193  
 SQ SEQUENCE 193 AA; 21180 MW; 413C8AB647B5C540 CRC64;

Query Match 85.7%; Score 48; DB 5; Length 193;  
 Best Local Similarity 70.0%; Pred. No. 0.52;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 CVIGYSGDXC 10  
 Db 56 CPLGYSGDYC 65  
 I :|||||:I

## RESULT 10

O8T523 PRELIMINARY; PRT; 194 AA.  
 ID O8T523  
 AC O8T523;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DE Glp-1 (Fragment).  
 OS Caenorhabditis briggsae.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6238;  
 RN [1]  
 RC SEQUENCE FROM N.A.

RP STRAIN=PB826;  
 RT Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;  
 "Levels of DNA polymorphism vary with mating system in the nematode  
 genus Caenorhabditis.";  
 RL Genetics 0:0-0(2002).  
 DR EMBL; AF491458; AAM09698.1; -.  
 FT NON\_TER 1 194  
 SQ SEQUENCE 194 AA; 21275 MW; 07313C8ABD9C65C5 CRC64;

Query Match 85.7%; Score 48; DB 5; Length 194;  
 Best Local Similarity 70.0%; Pred. No. 0.52;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 CVIGYSGDXC 10  
 Db 56 CPLGYSGDYC 65  
 I :|||||:I

## RESULT 11

O9GPM9 PRELIMINARY; PRT; 963 AA.  
 ID O9GPM9  
 AC O9GPM9;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DE Notch-like transmembrane receptor (Fragment).  
 OS Caenorhabditis briggsae.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6238;  
 RN [1]  
 RC SEQUENCE FROM N.A.

RP STRAIN=AF16;  
 RX MEDLINE=21100341; PubMed=11156985;  
 RA Rudel D., Kimble J.;  
 RT "Conservation of glp-1 regulation and function in nematodes.";  
 RL Genetics 157:639-654(2001).  
 DR EMBL; AF315555; AAG49317.1; -.  
 DR HSSP; P01132; IEFG.  
 DR InterPro; IPR002110; ANK.



```

DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR000800; Notch.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 4.
DR Pfam: PF00066; notch; 3.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00181; EGF; 5.
DR SMART; SM00001; EGF-like; 4.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00088; ANK_REPEAT; 3.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_4.
DR PROSITE; PS0186; EGF_2; 3.
DR ANK repeat; Receptor; Repeat; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 963 AA; 106394 MW; 52BF80010E12FB57 CRC64;

Query Match 85.7%; Score 48; DB 5; Length 963;
Best Local Similarity 70.0%; Pred. No. 2.7;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
| :|||||:|
DB 123 CPLYSGDYC 132

RESULT 12
Q8WTP0 PRELIMINARY; PRT; 1162 AA.
AC Q8WTP0;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Vitelline coat component HrvC120.
GN HrvC120.
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Pyuridae; Halocynthia.
OX NCBI_TaxID=7729;
RN [1]
RP SEQUENCE FROM N.A.
RA Sawada H., Sakai N., Abe Y., Tanaka E., Takahashi Y., Fujino J.,
RA Kodama E., Takizawa S., Yokosawa H.;
RT "A Novel Sperm Receptor VC70, Consisting of 12 EGF-like Repeats, Is
RT Degraded by the Ubiquitin-Proteasome Pathway during Ascidian
RT Fertilization."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB061740; BAB7021.1; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR001507; Endoglin/CD105.
DR Pfam: PF00008; EGF; 13.
DR PRINTS; PR00010; EGFBLLOOD.
DR SMART; SM00181; EGF; 14.
DR SMART; SM00241; ZP; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_13.
DR PROSITE; PS0186; EGF_2; UNKNOWN_10.
SQ SEQUENCE 1162 AA; 124499 MW; 26F57DE32B1A8E6C CRC64;

Query Match 85.7%; Score 48; DB 5; Length 1162;
Best Local Similarity 70.0%; Pred. No. 3.3;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
| :|||||:|

```

```

DB 650 CHVGSGDHC 659
| :|||||:|

RESULT 13
Q9GPN0 PRELIMINARY; PRT; 1270 AA.
AC Q9GPN0;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Notch-like transmembrane receptor.
GN GLP-1.
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF16;
RX MEDLINE=21100341; PubMed=11156985;
RA Rudel D., Kimble J.;
RT "Conservation of glp-1 regulation and function in nematodes.";
RL Genetics 157:639-654 (2001).
DR EMBL; AF315554; AAG49316.1; -.
DR HSSP; P01132; 1EGF.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR000800; Notch.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 10.
DR Pfam: PF00066; notch; 3.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00181; EGF; 12.
DR SMART; SM00179; EGF_Ca; 10.
DR SMART; SM00001; EGF-like; 8.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00088; ANK_REPEAT; 3.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_10.
DR PROSITE; PS0186; EGF_2; 8.
DR PROSITE; PS0187; EGF_Ca; 1.
DR ANK repeat; Receptor; Repeat; Transmembrane.
KW ANK repeat; Receptor; Repeat; Transmembrane.
SQ SEQUENCE 1270 AA; 138964 MW; A7662EB575A4B61B CRC64;

Query Match 85.7%; Score 48; DB 5; Length 1270;
Best Local Similarity 70.0%; Pred. No. 3.7;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
| :|||||:|
DB 430 CPLYSGDYC 439

RESULT 14
Q9WVB5 PRELIMINARY; PRT; 1531 AA.
AC Q9WVB5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SLIT1.
GN SLIT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER/ICR;
RX MEDLINE=99365246; PubMed=10433822;
RA Yuan W., Zhou L., Chen J.H., Wu J.Y., Rao Y., Ornitz D.M.;
RT "The mouse Slit family: secreted ligands for ROBO expressed in
RL Dev. Biol. 212:290-306(1999).
DR EMBL; AF144627; AA044758.1;
DR HSSP; P00743; ICCF
DR MGD; MGI:1315203; Slit1.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000359; Cys_knot.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003645; FcIn.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_Out.
DR InterPro; IPR003591; LRR_Typ.
DR Pfam; PF00008; EGF_9.
DR Pfam; PF00054; laminin_G; 1.
DR Pfam; PF00360; LRR; 19.
DR Pfam; PF01463; LRRCT; 4.
DR Pfam; PF01462; LRRNT; 4.
DR PRINTS; PR00010; EGFBLDOD.
DR SMART; SM00041; CT; 1.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF_Like; 7.
DR SMART; SM00274; FcIn; 2.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00370; LRR; 3.
DR SMART; SM00082; LRRCT; 4.
DR SMART; SM00013; LRRNT; 4.
DR SMART; SM00369; LRR_Typ; 10.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 2.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 1531 AA; 167546 MW; F7D09AA6693A4F30 CRC64;

Query Match 85.7% Score 48; DB 11; Length 1531;
Best Local Similarity 60.0% Pred. No. 4.4;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
DB 1071 CMLGTTGDC 1080

RESULT 15
Q9ROM0 ID Q9ROM0 PRELIMINARY; PRT; 2920 AA.
AC Q9ROM0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Flamingo 1.
GN CELSR2 OR FLAMINGO 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RX MEDLINE=99418630; PubMed=10490098;
RA Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RA Takeichi M., Uemura T.;
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
RT polarity under the control of frizzled.";
RL Cell 98:585-595(1999).
CC -1- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.
DR EMBL; AB028499; BAA84070.1;
DR HSSP; P00740; LEDM.
DR MGD; MGI:1858235; Celsr2.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR002049; Laminin_G.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00028; cadherin; 9.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRW; 1.
DR Pfam; PF00054; laminin_G; 2.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 9.
DR SMART; SM00180; EGF_Lam; 1.
DR SMART; SM00001; EGF_Like; 6.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormr; 1.
DR SMART; SM00282; LamG; 2.
DR SMART; SM00208; TNFR; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00232; CADHERIN_1; 5.
DR PROSITE; PS00268; CADHERIN_2; 9.
DR PROSITE; PS00022; EGF_1; UNKNOWN_6.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
KW Calcium-binding; Cell adhesion; EGF-like domain; Glycoprotein.
SQ SEQUENCE 2920 AA; 317648 MW; 2919558DF467114F CRC64;

Query Match 85.7% Score 48; DB 11; Length 2920;
Best Local Similarity 70.0% Pred. No. 8.6;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
DB 1820 CVLGYGDNC 1829
```

Search completed: July 2, 2003, 07:28:38  
Job time : 42.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 43.3333 seconds  
(without alignments)  
30.750 Million cell updates/sec

Title: US-09-673-785D-13  
Perfect score: 50  
Sequence: 1 CVTIGSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	10	21	AAV52143
2	50	100.0	47	23	AAG66047
3	50	100.0	53	5	AAP40315
4	50	100.0	53	10	AAP91658
5	50	100.0	53	11	AAR08007
6	50	100.0	53	16	AAR67275
7	50	100.0	53	18	AAW34466
8	50	100.0	53	18	AAW34467
9	50	100.0	53	19	AAW50134
10	50	100.0	53	19	AAW50139

11	50	100.0	53	19	AAW50140
12	50	100.0	53	22	AA37612
13	50	100.0	53	23	AAE15714
14	50	100.0	54	23	AAU76704
15	50	100.0	54	23	AAU76706
16	50	100.0	56	7	AAP60791
17	50	100.0	117	19	AAW68455
18	50	100.0	376	22	AAU02933
19	50	100.0	376	22	AAU02956
20	50	100.0	444	22	AAU02934
21	50	100.0	576	22	AAU02930
22	50	100.0	576	22	AAU02955
23	50	100.0	1073	22	AAU02953
24	50	100.0	1080	6	AAP50296
25	45	90.0	2146	22	AB62317
26	43	86.0	655	23	AAW52242
27	43	86.0	741	21	AAW34485
28	43	86.0	1473	23	AAE18208
29	43	86.0	2409	12	AAE12609
30	43	86.0	2469	23	AAE18207
31	43	86.0	2471	20	AAV06816
32	42	84.0	45	21	AAV76904
33	41	82.0	14	18	AAW34460
34	40	80.0	46	23	AAG66045
35	40	80.0	48	15	AAW57106
36	40	80.0	48	20	AAV01790
37	40	80.0	71	9	AAW81372
38	40	80.0	73	9	AAW81367
39	40	80.0	298	23	AAU11636
40	40	80.0	308	22	AAW35196
41	40	80.0	366	22	AAW35195
42	40	80.0	366	22	AAW35202
43	40	80.0	601	23	ABW93437
44	40	80.0	754	18	AAW27536
45	40	80.0	1296	23	ABG66702

#### ALIGNMENTS

#### RESULT 1

AAV52143  
ID AAV52143 standard; peptide: 10 AA.

XX AAV52143;

XX 28-JAN-2000 (first entry)

DT Mouse EGF derived peptide for targeting laminin receptor.

DE Epidermal growth factor; EGF; laminin receptor; angiogenesis;  
KW medicament; wound healing; retinopathy of immaturity; metastatic cancer;  
KW candida infection; leishmania; trichomonas vaginalis.

XX Mus sp.

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX WPI; 2000-013229/01.  
 XX New peptide derived from murine epidermal growth factor (MEGF)  
 XX Claim 4; Page 28; 35pp; English.  
 XX This is a peptide derived from mouse epidermal growth factor (EGF)  
 XX residues 33-42. This peptide is used in the invention to prepare a  
 XX composition to target laminin receptors. EGF derived peptides inhibit  
 XX blood vessel formation through their antagonism of the high affinity  
 XX laminin receptor found on endothelial cells. The peptide is modified  
 XX from the natural sequence to prevent protease attack. The peptide is used  
 XX in the preparation of a medicament for binding to laminin receptors as an  
 XX (ant)agonist. The medicament is also useful for healing endothelial cell  
 XX wounds and treating angiogenic diseases, especially retinopathy of  
 XX immaturity. Other diseases treated include metastatic cancer,  
 XX Candida spp. infection, and parasitic infestations like leishmania and  
 XX trichomonas vaginalis. The peptide are anti-angiogenic in human models.  
 XX The peptides also inhibit both laminin- and EGF-stimulated angiogenesis,  
 XX and prevent tumour cell attachment to basement membranes.  
 XX Sequence 10 AA;  
 Query Match 100.0%; Score 50; DB 21; Length 10;  
 Best Local Similarity 80.0%; Pred. No. 0.51;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXGDXC 10  
 ||||:||||:  
 Db 1 CVIGYGDRC 10  
 RESULT 2  
 ID AAG66047  
 XX AAG66047 standard; peptide; 47 AA.  
 AC AAG66047;  
 XX 27-FEB-2002 (first entry)  
 DT Mouse EGF motif sequence.  
 DE ErBB-4; neuregulin-4; NRG-4; pro-NRG-4; neuroprotective; vulnerary;  
 KW cerebroprotective; vasotropic; antiparkinsonian; anticonvulsant;  
 KW cytotstatic; nootropic; EGF; NRG-2.  
 OS Mus musculus.  
 OS WO200181540-A2.  
 PN 01-NOV-2001.  
 PD 20-APR-2001; 2001WO-IL00371.  
 PF 21-APR-2000; 2000US-0553769.  
 PR (YEDA ) YEDA RES & DEV CO LTD.  
 XX Harari D, Yarden Y;  
 PI WPI; 2002-041398/05.  
 DR Novel ErBB-4 ligand, referred as neuregulin (NRG)-4 and polynucleotide  
 XX sequences encoding NRG-4, useful for upregulating or downregulating  
 PT ErBB-4 receptor activity to treat Alzheimer's disease, stroke, gastric  
 PT cancer  
 XX Disclosure; Fig 1c; 153pp; English.  
 XX The invention relates to a novel ErBB-4 ligand, neuregulin-4 (NRG-4).  
 CC NRG-4 binds to mammalian ErBB-4 receptor and can be expressed by standard  
 CC recombinant methodology. Pharmaceutical compositions comprising NRG-4 are

CC useful for regulating an endogenous protein affecting ErBB-4 receptor  
 CC activity in vivo. They are also useful for treating or preventing a  
 CC disease condition or syndrome associated with dysregulation of an  
 CC endogenous protein affecting ErBB-4 receptor activity, e.g., amyotrophic  
 CC lateral sclerosis (Lou Gehrig's disease), Bell's palsy, spinal muscular  
 CC atrophy, brain trauma, stroke, ischemia, Alzheimer's disease, Parkinson's  
 CC disease, epilepsy, multiple sclerosis, Huntington's chorea, Down's  
 CC syndrome, nerve deafness, neuropathy, muscular dystrophy, extramammary  
 CC Paget's disease, gastric, pancreatic, prostate, breast and ovarian  
 CC cancer, cervical carcinoma, endometrial adenocarcinoma, pancreatic D  
 CC cells-somatostatinoma and Zollinger-Ellison syndrome. The agent comprised  
 CC in the pharmaceutical composition includes a polypeptide (e.g., a soluble  
 CC ligand binding domain of ErBB-4 i.e., IgB4; or a monoclonal, polyclonal,  
 CC humanized, single chain antibody or an immunoreactive derivative of an  
 CC antibody) capable of binding the endogenous protein affecting ErBB-4  
 CC receptor activity. Traceable synthetic/recombinant NRG-4-tagged molecules  
 CC can serve as a diagnostic tool in which cells binding NRG-4 can be  
 CC measured. Sequences AAG66044-53 represent the EGF-like motifs of various  
 CC growth factors.  
 XX Sequence 47 AA;  
 Query Match 100.0%; Score 50; DB 23; Length 47;  
 Best Local Similarity 80.0%; Pred. No. 2;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXGDXC 10  
 ||||:||||:  
 Db 33 CVIGYGDRC 42  
 RESULT 3  
 ID AAP40315  
 XX AAP40315 standard; protein; 53 AA.  
 AC AAP40315;  
 XX 22-JUL-1992 (first entry)  
 DT Epidermal growth factor tripentacontapeptide.  
 DE EGF.  
 KW Synthetic.  
 OS Key Location/Qualifiers  
 FH Disulfide-bond 6..20  
 FT Disulfide-bond 14...31  
 FT Disulfide-bond 33...42  
 XX JPS9027858-A.  
 PN 14-FEB-1984.  
 PD 05-AUG-1982; 82JP-0137128.  
 PF 05-AUG-1982; 82JP-0137128.  
 PR (NNSH ) NIPPON SHINYAKU KK.  
 PA WPI; 1984-072465/12.  
 DR  
 XX Synthesis of epidermal growth factor polypeptide - by condensation  
 PT of protected smaller peptide sequences, de:protection then oxidn. to  
 PT cyclise.  
 XX Claim1; Page 1; 8pp; Japanese.

XX The amino acid sequence is that of an epidermal growth factor  
 CC tripentacontapeptide which is synthesised by condensation of  
 CC protected smaller peptide sequences. This method produces the  
 CC peptide smoothly, with high purity and yield.

SQ Sequence 53 AA;  
 Query Match 100.0%; Score 50; DB 5; Length 53;  
 Best Local Similarity 80.0%; Pred. No. 2.3;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
 ||||:||||  
 DB 33 CVIGYSGDRC 42

RESULT 4  
 AAP91658  
 ID AAP91658 standard; protein; 53 AA.  
 AC AAP91658;  
 XX  
 DT 29-JUN-1990 (first entry)  
 XX  
 DE Amino acid sequence for naturally occurring epidermal growth factor (EGF).  
 DE  
 XX  
 KW Epidermal growth factor; angiogenesis; synthetic peptide.  
 XX  
 PN WO8901489-A.  
 XX  
 XX  
 PD 23-FEB-1989.  
 XX  
 XX 10-AUG-1988; 88WO-AU00300.  
 XX  
 PR 10-AUG-1987; 87AU-0003629.  
 XX  
 XX (CSIR ) COMMONWEALTH SCIENT ORG.  
 XX  
 PI McAuslan BR;  
 XX  
 DR WPI; 1989-068852/09.  
 XX  
 XX Synthetic peptide active in stimulating angiogenesis -  
 PT has sequences corresponding to amino acid sequences occurring in  
 PT epidermal growth factor.  
 XX  
 PS Fig 1; 1/1; 1lpp; English.  
 XX  
 CC The inventors claim synthetic peptides which correspond to sequences  
 CC occurring in EGF, but excluding EGF. The synthetic peptides  
 CC correspond to active sequences from EGF fragments: 3-14, 12-14, 12-15,  
 CC 3-10, 29-37 and 33-37 (Claims 2-7 and AAP91659-p91664 resp.). The  
 CC peptides are angiogenic. Their relative shortness means that they pose  
 CC fewer synthesis problems than the entire EGF molecule.  
 XX  
 SQ Sequence 53 AA;  
 Query Match 100.0%; Score 50; DB 10; Length 53;  
 Best Local Similarity 80.0%; Pred. No. 2.3;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
 ||||:||||  
 DB 33 CVIGYSGDRC 42

RESULT 5  
 AAR08007  
 ID AAR08007 standard; protein; 53 AA.  
 XX  
 AC AAR08007;  
 XX  
 DT 25-FEB-1991 (first entry)  
 XX  
 DE Modified murine epidermal growth factor.  
 XX  
 KW Modified murine epidermal growth factor; stability; storage;

KW epithelial wounds; gastric acid secretion.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 11..11  
 FT /label= E, N, Q, A, K  
 XX  
 PN WO9013570-A.  
 XX  
 PD 15-NOV-1990.  
 XX  
 PF 09-MAY-1990; 90WO-US02600.  
 XX  
 PR 12-MAY-1989; 89US-0351773.  
 XX  
 PA (CHIR-) CHIRON CORP.  
 XX  
 XX Nascimento CG, Medina-Selby A;  
 PI WPI; 1990-361427/48.  
 DR  
 XX Human epidermal growth factor - is substituted at position 11 for  
 PT greater stability and improved storage life.  
 PT  
 XX Claim 9; Page 25; 32pp; English.  
 PS  
 XX The human EGF is used to treat oversecretion of gastric acid or an  
 CC epithelial wound. EGF is modified to increase its chemical  
 CC stability. Its storage life is improved without diminishing its  
 CC biological activity. The proteins may be prepared by traditional  
 CC chemical or recombinant means.  
 CC See also AAR08004.  
 XX  
 SQ Sequence 53 AA;  
 Query Match 100.0%; Score 50; DB 11; Length 53;  
 Best Local Similarity 80.0%; Pred. No. 2.3;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
 ||||:||||  
 DB 33 CVIGYSGDRC 42

RESULT 6  
 AAR67275  
 ID AAR67275 standard; peptide; 53 AA.  
 XX  
 AC AAR67275;  
 XX  
 DT 01-AUG-1995 (first entry)  
 XX  
 DE Synthetic human/mouse EGF.  
 XX  
 KW Cell growth factor; viscoelastic solution; fibroblastic growth factor;  
 KW FGF; epidermal growth factor; EGF; buffered solution; lubrication;  
 KW carboxypropyl methylcellulose; cellulose gum; dextran; dextran sulphate;  
 KW chondroitin sulphate; sodium hyaluronate; osmolality; mitogenic;  
 KW wound healing; cell protection; cell coating; surgery; tissue space;  
 KW hydroxypropyl methylcellulose; manipulation.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 37  
 FT /note= "Given in the specification as Try"  
 XX  
 PN US5366964-A.  
 XX  
 PD 22-NOV-1994.  
 XX  
 XX 15-DEC-1988; 88US-0284533.  
 PF  
 XX 15-DEC-1988; 88US-0284533.  
 PR

PR 13-NOV-1989; 89US-0434305.  
 XX (LIND/) LINDSTROM R L.  
 PA (SKEL/) SKELNIK D.  
 XX  
 PI Lindstrom RL, Skelnik D;  
 XX WPI; 1995-005865/01.  
 DR

XX Viscoclastic soln. contg. hydroxypropyl methylcellulose, sodium  
 PT hyaluronate, chondroitin sulphate and growth factors - used as a  
 PT surgical soln. to promote wound healing, e.g. of corneal cells  
 XX  
 PS Disclosure; Column 5; 8pp; English.  
 XX

CC The sequences given in AAR67273-76 are cell growth factors which may  
 CC be used in the viscoelastic solution of the invention. The peptides  
 CC are derived from fibroblastic growth factor (FGF) and epidermal  
 CC growth factor (EGF). The viscoelastic solution also comprises a  
 CC buffered solution which is pref. a buffered balanced salt solution,  
 CC at least 1 of hydroxy- or carboxypropyl methylcellulose, a cellulose  
 CC gum, dextran or dextran sulphate, chondroitin sulphate, and sodium  
 CC hyaluronate. It has a pH of 6-8 and an osmolarity of 200-400 mOsmol/l.  
 CC The growth factors are mitogenic in vitro for a wide range of tissues  
 CC and the viscoelastic soln. may be used as a surgical soln. which is  
 CC in direct contact with cells undergoing wound healing. It also provides  
 CC a cell protection and cell coating during surgery. The soln. provides  
 CC maintenance of tissue space, hydroxypropyl methylcellulose and  
 CC chondroitin sulphate lubricate the tissue, while sodium hyaluronate  
 CC provides tissue manipulation.  
 XX

SQ Sequence 53 AA;

Query Match 100.0%; Score 50; DB 16; Length 53;  
 Best Local Similarity 80.0%; Pred. No. 2.3;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDXC 10  
 ||||:||||:  
 Db 33 CVIGYSGDRC 42

RESULT 7  
 AAW34466  
 ID AAW34466 standard; Protein; 53 AA.  
 XX  
 AC AAW34466;  
 XX

DT 08-MAY-1998 (first entry)

XX Chimeric epidermal growth factor protein AC.

DE Epidermal growth factor; EGF; chimeric; human; mouse; recombinant.

KW Chimeric - Mus sp.

XX Chimeric - Homo sapiens.

OS Key Location/Qualifiers

FT Region 1..10 /note= "region A from mouse EGF"  
 FT Region 11..33 /note= "region B from human EGF"  
 FT Region 34..47 /note= "region C from mouse EGF"  
 FT Region 48..53 /note= "region D from human EGF"  
 XX

XX AU9717759-A.

PN 30-OCT-1997.

XX 08-APR-1997; 97AU-0017759.  
 PF

XX

PR 23-APR-1996; 96JP-0123970.  
 XX (HGET ) HIGETA SHOYU KK.  
 PA

XX Tagami H, Tanaka A;  
 PI

XX WPI; 1997-550187/51.  
 DR N-PSDB; AAT99941.

XX Chimeric epidermal growth factor proteins - and DNA molecules for  
 PT their recombinant production  
 PT

XX Claim 5; Page 26; 40pp; English.

XX This is a chimeric epidermal growth factor (EGF) protein AC. This  
 CC chimeric EGF protein contains sequences derived from human and mouse  
 CC EGF sequences. The chimeric EGF proteins can be recombinantly produced  
 CC by culturing a microorganism of the genus *Bacillus* transformed by an  
 CC expression vector containing the encoding DNA molecules. The recombinant  
 CC chimeric EGF proteins exhibit a wide variety of physiological activities  
 CC similar to those of naturally occurring EGF.  
 XX

SQ Sequence 53 AA;

Query Match 100.0%; Score 50; DB 18; Length 53;  
 Best Local Similarity 80.0%; Pred. No. 2.3;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDXC 10  
 ||||:||||:  
 Db 33 CVIGYSGDRC 42

RESULT 8  
 AAW34467  
 ID AAW34467 standard; Protein; 53 AA.  
 XX

AC AAW34467;  
 XX

DT 08-MAY-1998 (first entry)

XX Chimeric epidermal growth factor protein ABC.

DE Epidermal growth factor; EGF; chimeric; human; mouse; recombinant.

KW Chimeric - Mus sp.

XX Chimeric - Homo sapiens.

OS Key Location/Qualifiers

FT Region 1..10 /note= "region A from mouse EGF"  
 FT Region 11..33 /note= "region B from mouse EGF"  
 FT Region 34..47 /note= "region C from mouse EGF"  
 FT Region 48..53 /note= "region D from human EGF"  
 XX

XX AU9717759-A.

PN 30-OCT-1997.

XX 08-APR-1997; 97AU-0017759.  
 PF

XX 23-APR-1996; 96JP-0123970.  
 PR

XX (HGET ) HIGETA SHOYU KK.  
 PA

XX Tagami H, Tanaka A;  
 PI

XX WPI; 1997-550187/51.  
 DR N-PSDB; AAT99942.

XX

us-09-673-785d-13.jull.rag

Wed Jul 2, 08:08:16 2003

PT Chimeric epidermal growth factor proteins - and DNA molecules for  
 their recombinant production  
 XX  
 PS Claim 6; Page 26; 40pp; English.  
 XX This is a chimeric epidermal growth factor (EGF) protein ABC. This  
 CC chimeric EGF protein contains sequences derived from human and mouse  
 CC EGF sequences. The chimeric EGF proteins can be recombinantly produced  
 CC by culturing a microorganism of the genus *Bacillus* transformed by an  
 CC expression vector containing the encoding DNA molecules. The recombinant  
 CC chimeric EGF proteins exhibit a wide variety of physiological activities  
 CC similar to those of naturally occurring EGF.  
 XX  
 SQ Sequence 53 AA;  
 Query Match 100.0%; Score 50; DB 18; Length 53;  
 Best Local Similarity 80.0%; Pred. No. 2.3;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDNC 10  
 Db 33 CVIGYSGDRC 42  
 RESULT 9  
 AAW50134  
 ID AAW50134 standard; protein; 53 AA.  
 AC AAW50134;  
 XX  
 DT 08-JUL-1998 (first entry)  
 DE Mouse epidermal growth factor.  
 XX  
 KW Mouse; epidermal growth factor; EGF; isolation; recombinant;  
 KW ion exchange chromatography; *Bacillus brevis*.  
 XX  
 OS Mus sp.  
 XX  
 PN AU9728698-A.  
 XX  
 PD 05-FEB-1998.  
 XX  
 PF 17-JUL-1997; 97AU-0028698.  
 XX  
 PR 01-AUG-1996; 96JP-0218109.  
 XX  
 PA (HGET ) HIGETA SHOYU KK.  
 XX  
 PI Miyauchi A, Nemoto A, Warren B;  
 XX  
 DR WPI; 1998-131057/13.  
 XX  
 PT Isolation of recombinant epidermal growth factor from whole broth -  
 PT by fluidised-bed ion-exchange chromatography  
 XX  
 PS Claim 6; Page 26; 38pp; English.  
 XX  
 CC The present sequence represents mouse epidermal growth factor (EGF)  
 CC from the present invention. The present invention describes a method for  
 CC the isolation of recombinant epidermal growth factor (EGF) from whole  
 CC broth. The method comprises: (a) passing a culture containing the  
 CC protein upwards through a column containing a fluidised bed of ion  
 CC exchanger, washing the ion exchanger to remove non-adsorbed material,  
 CC allowing an eluant downwards through the column; (b) filtering the eluate  
 CC through a membrane with a molecular weight cutoff of 10000, and (c)  
 CC concentrating the filtrate on a membrane with a molecular weight cutoff  
 CC of 5000. The method is for isolating recombinant human, mouse, pig or  
 CC rat EGF, or an EGF analogue with one of four defined amino acid sequences  
 CC (given in the specification), from *Bacillus brevis* cultures. The EGF  
 CC proteins can be recovered in high yield and high purity without the need  
 CC to pretreat the culture broth to remove the cells, e.g. by costly

CC centrifugation or membrane filtration.

XX  
 SQ Sequence 53 AA;  
 Query Match 100.0%; Score 50; DB 19; Length 53;  
 Best Local Similarity 80.0%; Pred. No. 2.3;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDNC 10  
 Db 33 CVIGYSGDRC 42  
 RESULT 10  
 AAW50139  
 ID AAW50139 standard; protein; 53 AA.  
 AC AAW50139;  
 XX  
 DT 08-JUL-1998 (first entry)  
 DE Epidermal growth factor analogue 3.  
 XX  
 KW Epidermal growth factor; EGF; isolation; recombinant;  
 KW ion exchange chromatography; *Bacillus brevis*.  
 XX  
 OS Unidentified.  
 XX  
 PN AU9728698-A.  
 XX  
 PD 05-FEB-1998.  
 XX  
 PF 17-JUL-1997; 97AU-0028698.  
 XX  
 PR 01-AUG-1996; 96JP-0218109.  
 XX  
 PA (HGET ) HIGETA SHOYU KK.  
 XX  
 PI Miyauchi A, Nemoto A, Warren B;  
 XX  
 DR WPI; 1998-131057/13.  
 XX  
 PT Isolation of recombinant epidermal growth factor from whole broth -  
 PT by fluidised-bed ion-exchange chromatography  
 XX  
 PS Claim 6; Page 28; 38pp; English.  
 XX  
 CC The present sequence represents a protein which has epidermal growth  
 CC factor (EGF) like activity, from the present invention. The present  
 CC invention describes a method for the isolation of recombinant epidermal  
 CC growth factor (EGF) from whole broth. The method comprises: (a) passing  
 CC a culture containing the protein upwards through a column containing a  
 CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-  
 CC adsorbed material, allowing the ion exchanger to settle, and eluting  
 CC adsorbed protein by passing an eluant downwards through the column; (b)  
 CC filtering the eluate through a membrane with a molecular weight cutoff  
 CC of 10000, and (c) concentrating the filtrate on a membrane with a  
 CC molecular weight cutoff of 5000. The method is for isolating recombinant  
 CC human, mouse, pig or rat EGF, or an EGF analogue with one of four  
 CC defined amino acid sequences (given in the specification), from *Bacillus*  
 CC *brevis* cultures. The EGF proteins can be recovered in high yield and  
 CC high purity without the need to pretreat the culture broth to remove the  
 CC cells, e.g. by costly centrifugation or membrane filtration.  
 XX  
 SQ Sequence 53 AA;  
 Query Match 100.0%; Score 50; DB 19; Length 53;  
 Best Local Similarity 80.0%; Pred. No. 2.3;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDNC 10  
 Db 33 CVIGYSGDRC 42

RESULT 11  
AAW50140  
ID AAW50140 standard; protein; 53 AA.  
XX AC AAW50140;  
XX DT 08-JUL-1998 (first entry)  
XX DE Epidermal growth factor analogue 4.  
XX DE Epidermal growth factor; EGF; isolation; recombinant;  
KW ion exchange chromatography; Bacillus brevis.  
XX OS Unidentified.  
XX PN AU9728698-A.  
XX PD 05-FEB-1998.  
XX PF 17-JUL-1997; 97AU-0028698.  
XX PR 01-AUG-1996; 96JP-0218109.  
XX PA (HGET.) HIGETA SHOYU KK.  
XX PI Miyauchi A, Nemoto A, Warren B;  
XX WPI; 1998-131057/13.  
XX PT Isolation of recombinant epidermal growth factor from whole broth -  
PT by fluidised-bed ion-exchange chromatography  
XX PS Claim 6; Page 29; 38pp; English.  
XX CC The present sequence represents a protein which has epidermal growth  
CC factor (EGF) like activity, from the present invention. The present  
CC invention describes a method for the isolation of recombinant epidermal  
CC growth factor (EGF) from whole broth. The method comprises: (a) passing  
CC a culture containing the protein upwards through a column containing a  
CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-  
CC adsorbed material, allowing the ion exchanger to settle, and eluting  
CC adsorbed protein by passing an eluant downwards through the column; (b)  
CC filtering the eluate through a membrane with a molecular weight cutoff  
CC of 10000, and (c) concentrating the filtrate on a membrane with a  
CC molecular weight cutoff of 5000. The method is for isolating recombinant  
CC human, mouse, pig or rat EGF, or an EGF analogue with one of four  
CC defined amino acid sequences (given in the specification), from Bacillus  
CC brevis cultures. The EGF proteins can be recovered in high yield and  
CC high purity without the need to pretreat the culture broth to remove the  
CC cells, e.g. by costly centrifugation or membrane filtration.  
XX SQ Sequence 53 AA;  
Query Match 100.0%; Score 50; DB 19; Length 53;  
Best Local Similarity 80.0%; Pred. NO. 2.3;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CVIGXSGDNC 10  
Db 33 CVIGYSGDRC 42  
RESULT 12  
AAB37612  
ID AAB37612 standard; protein; 53 AA.  
XX AC AAB37612;  
XX DT 27-FEB-2001 (first entry)  
XX DE Human EGF.  
XX

XX KW Merozoite surface protein; protazoacide; vaccine; malaria; human; EGF;  
KW epidermal growth factor.  
XX OS Homo sapiens.  
XX PN WO200063245-A2.  
XX PD 26-OCT-2000.  
XX PF 20-APR-2000; 2000WO-GB01558.  
XX PR 20-APR-1999; 99GB-0009072.  
XX PR 13-MAY-1999; 99US-0311817.  
XX PR 25-MAY-1999; 99CA-2271451.  
XX XX (MEDI-) MEDICAL RES COUNCIL.  
XX PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;  
XX WPI; 2001-015762/02.  
XX DR Novel variants of the C-terminal fragment of Plasmodium merozoite  
XX PT surface protein-1, useful as vaccines for treating or preventing  
XX PT malaria -  
XX PS Disclosure; Fig 1; 126pp; English.  
XX CC The present invention relates to non-natural variants of a C-terminal  
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The  
CC non-natural variants have reduced affinity for at least 1 antibody  
CC capable of blocking a second antibody that inhibits the proteolytic  
CC cleavage of Plasmodium MSP-1<sub>4-2</sub>, and has the same affinity for at least  
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium  
CC MSP-1<sub>4-2</sub>, compared to natural MSP-1<sub>1-9</sub>. The non-natural variants of  
CC the present invention are useful for immunising a mammal against malaria,  
CC and can be used to treat malaria. The present sequence is human  
CC epidermal growth factor (EGF). This sequence was used in a sequence  
CC homology comparison with the wild-type MSP-1 protein from P. falciparum  
CC (see AAB37608), which was used to generate the variants of the present  
CC invention.  
XX SQ Sequence 53 AA;  
Query Match 100.0%; Score 50; DB 22; Length 53;  
Best Local Similarity 80.0%; Pred. NO. 2.3;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CVIGXSGDNC 10  
Db 33 CVIGYSGDRC 42  
RESULT 13  
AAE15714  
ID AAE15714 standard; Protein; 53 AA.  
XX AC AAE15714;  
XX DT 12-MAR-2002 (first entry)  
XX DE Epidermal growth factor (EGF) used in exemplification of the invention.  
XX KW Humanised form; monoclonal antibody alpha 340; gene therapy;  
KW epidermal growth factor receptor; EGF; cancer; colorectal; lung; breast;  
KW gastric; ovarian; immune response; cytostatic; cell growth; apoptosis;  
XX inhibitor.  
XX OS Unidentified.  
XX PN WO20018138-A1.  
XX PD 22-NOV-2001.



XX PF 21-MAY-2001; 2001WO-GB02226.  
 XX PR 19-MAY-2000; 2000GB-0011981.  
 XX PR 24-AUG-2000; 2000GB-0020794.  
 XX PA (SCAN-) SCANCELL LTD.  
 XX XX Ellis JRM, Durrant LG;  
 XX PI WPI; 2002-062384/08.  
 XX DR  
 XX PT New humanized form of mouse monoclonal antibody 340 which binds to  
 XX PT epidermal growth factor receptor and inhibits binding of growth factor,  
 XX PT useful for treating colorectal, lung, breast, gastric and ovarian  
 XX PT cancer -  
 XX PS Example 5; Fig 12; 53pp; English.  
 XX XX  
 XX The present invention relates to a humanised form of the antibody 340 (a  
 XX mouse monoclonal antibody which binds to epidermal growth factor (EGF)  
 XX receptor and inhibits binding of EGF), obtainable from the cell line  
 XX deposited with the ECACC under accession number 97021428. The humanised  
 XX form of the antibody 340 is useful in gene therapy, medicine and in the  
 XX manufacture of a medicament for treatment or prophylaxis of cancer. The  
 XX invention is useful for treating colorectal, lung, breast, gastric or  
 XX ovarian cancers or also for preventing the recurrence of cancer after  
 XX initial treatment or surgery. The invention is also useful for enhancing  
 XX a protective immune response against cancer by optimised immunisation  
 XX schedules. The humanised form of the antibody 340 has reduced  
 XX immunogenicity but shows similar binding to cells expressing EGF  
 XX receptor, as the original murine antibody and has increased ability to  
 XX inhibit the growth of EGF receptor expressing cells. The invention is  
 XX used as cell growth and apoptosis inhibitor. The present sequence  
 XX is epidermal growth factor (EGF) which is used in the exemplification of  
 XX the invention.  
 XX SQ Sequence 53 AA;  
 Query Match 100.0%; Score 50; DB 23; Length 53;  
 Best Local Similarity 80.0%; Pred. No. 2.3;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDXC 10  
 ||||:||||:  
 Db 33 CVIGYSGDRC 42  
 RESULT 14  
 AAU76704  
 ID AAU76704 standard; Protein; 54 AA.  
 XX AC AAU76704;  
 XX XX  
 XX DT 21-MAY-2002 (first entry)  
 XX DE Mouse epidermal growth factor (EGF).  
 XX KW Mouse; epidermal growth factor; receptor; EGF;  
 XX KW chimeric polynucleotide; directed evolution; chimeraogenesis;  
 XX KW recombination.  
 XX OS Mus sp.  
 XX XX  
 XX FH Key Location/Qualifiers  
 XX FT Misc-difference 54  
 XX FT /label= Unknown  
 XX FT /note= "Encoded by TAA"  
 XX XX  
 XX PN WO200206469-A2.  
 XX PD 24-JAN-2002.  
 XX XX

PF 18-JUL-2001; 2001WO-US22640.  
 XX PR 18-JUL-2000; 2000US-218921P.  
 XX PR 18-JUL-2000; 2000US-219085P.  
 XX PR 19-OCT-2000; 2000US-0691873.  
 XX PR 19-OCT-2000; 2000US-0692732.  
 XX PA (ENCH-) ENCHIRA BIOTECHNOLOGY CORP.  
 XX XX  
 XX PI Coco WM, Encell LP, Arensdorf JJ;  
 XX DR WPI; 2002-179795/23.  
 XX DR N-PSDB; ABK10313.  
 XX XX  
 XX PT Forming chimeric polynucleotide by contacting single-stranded (ss)  
 XX PT scaffold fragment, and donor fragment populations to form hybridised  
 XX PT complex having ss regions which are filled-in, and ligating adjacent  
 XX PT fragments -  
 XX PS Example 2; Fig 3; 65pp; English.  
 XX XX  
 XX The invention describes a method of forming (M1) chimeric  
 XX polynucleotides (I). The method comprises contacting a population of  
 XX single-stranded scaffold fragments (SF) with population of donor  
 XX fragments (DF) to form a complex (II) comprising at least one SF  
 XX hybridised to two DFs. (II) is treated so that single stranded regions of  
 XX (II) are filled-in and adjacent fragments are ligated. (M1) is useful for  
 XX a directed evolution process which involves forming a library of (I)  
 XX that can be screened for a characteristic of interest. Subsequent rounds  
 XX of directed evolution can produce chimeric polynucleotides with an  
 XX improved characteristic of interest. The methods facilitate the  
 XX generation of chimeric polynucleotides and do not require hybridising  
 XX donor fragments to a target- or full-length template. Because the  
 XX chimeraogenesis process does not rely upon a contiguous, full-length  
 XX template, it is unnecessary to modify a template to facilitate its  
 XX removal. This is the amino acid sequence of the mouse epidermal growth  
 XX factor (EGF) used to demonstrate a method of in vitro recombination  
 XX described in the method of the invention.  
 XX SQ Sequence 54 AA;  
 Query Match 100.0%; Score 50; DB 23; Length 54;  
 Best Local Similarity 80.0%; Pred. No. 2.3;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDXC 10  
 ||||:||||:  
 Db 33 CVIGYSGDRC 42  
 RESULT 15  
 AAU76706  
 ID AAU76706 standard; Protein; 54 AA.  
 XX AC AAU76706;  
 XX XX  
 XX DT 21-MAY-2002 (first entry)  
 XX DE Modified mouse epidermal growth factor (EGF).  
 XX KW Mouse; epidermal growth factor; receptor; EGF;  
 XX KW chimeric polynucleotide; directed evolution; chimeraogenesis;  
 XX KW recombination.  
 XX OS Mus sp.  
 XX OS Synthetic.  
 XX XX  
 XX FH Key Location/Qualifiers  
 XX FT Misc-difference 38  
 XX FT /note= "Encoded by ACC"  
 XX PN WO200206469-A2.  
 XX XX

PD 24-JAN-2002.  
 XX  
 PF 18-JUL-2001; 2001WO-US22640.  
 XX  
 PR 18-JUL-2000; 2000US-218921P.  
 PR 18-JUL-2000; 2000US-219085P.  
 PR 19-OCT-2000; 2000US-0691873.  
 PR 19-OCT-2000; 2000US-0692732.  
 XX  
 PA (ENCH-) ENCHIRA BIOTECHNOLOGY CORP.  
 XX  
 XX Cocco WM, Ensell LP, Arensdorf JJ;  
 PI WPI: 2002-179795/23.  
 XX N-PSDB; ABK10315.  
 DR  
 DR  
 XX  
 PT Forming chimeric polynucleotide by contacting single-stranded (ss)  
 PT scaffold fragment, and donor fragment populations to form hybridised  
 PT complex having ss regions which are filled-in, and ligating adjacent  
 PT fragments  
 XX  
 PS Example 2; Fig 3; 65pp; English.  
 XX  
 CC The invention describes a method of forming (M1) chimeric  
 CC polynucleotides (I). The method comprises contacting a population of  
 CC single-stranded scaffold fragments (SF) with population of donor  
 CC fragments (DF) to form a complex (II) comprising at least one SF  
 CC hybridised to two Dfs. (II) is treated so that single stranded regions of  
 CC (II) are filled-in and adjacent fragments are ligated. (M1) is useful for  
 CC a directed evolution process which involves forming a library of (I)  
 CC that can be screened for a characteristic of interest. Subsequent rounds  
 CC of directed evolution can produce chimeric polynucleotides with an  
 CC improved characteristic of interest. The methods facilitate the  
 CC generation of chimeric polynucleotides and do not require hybridising  
 CC donor fragments to a target- or full-length template. Because the  
 CC chimeragenesis process does not rely upon a contiguous, full-length  
 CC template, it is unnecessary to modify a template to facilitate its  
 CC removal. This is the amino acid sequence of a mouse epidermal growth  
 CC factor (EGF) modified to make it as similar as possible to the human  
 CC protein (AAU76705) and used to demonstrate a method of in vitro  
 CC recombination described in the method of the invention.  
 XX  
 SQ Sequence 54 AA;  
 Query Match 100.0%; Score 50; DB 23; Length 54;  
 Best Local Similarity 80.0%; Pred. No. 2, 3;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDXC 10  
 Db 33 CVIGYSGDRC 42  
 ||||:||||:|

Search completed: July 2, 2003, 07:30:55  
 Job time : 43.3333 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 10.6667 Seconds  
(without alignments)  
27.584 Million cell updates/sec

Title: US-09-673-785D-13  
Perfect score: 50  
Sequence: 1 CVIGXSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/FCIUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	12	1 US-08-597-545-7	Sequence 7, Appl
2	50	100.0	12	1 US-08-457-135-7	Sequence 7, Appl
3	50	100.0	53	1 US-08-284-923-2	Sequence 2, Appl
4	50	100.0	53	4 US-08-619-032B-2	Sequence 2, Appl
5	50	100.0	53	6 5332669-2	Patent No. 5332669
6	43	86.0	2409	6 5180808-2	Patent No. 5180808
7	43	86.0	2471	1 US-08-185-432-16	Sequence 16, Appl
8	43	86.0	2471	1 US-08-083-590A-19	Sequence 19, Appl
9	43	86.0	2471	3 US-08-532-384-19	Sequence 19, Appl
10	43	86.0	2471	4 US-08-899-232-1	Sequence 1, Appl
11	42	84.0	154	3 US-09-191-647-10	Sequence 10, Appl
12	42	84.0	154	4 US-09-540-245A-10	Sequence 10, Appl
13	42	84.0	154	4 US-09-540-153-10	Sequence 10, Appl
14	40	80.0	48	6 5332669-1	Patent No. 5332669
15	40	80.0	754	2 US-08-525-864A-2	Sequence 2, Appl
16	39	78.0	17	6 5183805-1	Patent No. 5183805
17	39	78.0	17	6 5183805-2	Patent No. 5183805
18	39	78.0	25	6 5256643-8	Patent No. 5256643
19	39	78.0	37	2 US-08-039-364-14	Sequence 14, Appl
20	39	78.0	37	4 US-09-158-710-14	Sequence 14, Appl
21	39	78.0	42	1 US-08-168-091A-44	Sequence 44, Appl
22	39	78.0	44	1 US-08-278-089A-24	Sequence 24, Appl
23	39	78.0	46	3 US-08-899-437-12	Sequence 12, Appl
24	39	78.0	46	4 US-09-126-121-12	Sequence 12, Appl
25	39	78.0	47	3 US-08-753-007A-17	Sequence 17, Appl
26	39	78.0	47	4 US-09-398-496-17	Sequence 17, Appl
27	39	78.0	48	4 US-09-020-880-15	Sequence 15, Appl

28 39 78.0 48 4 US-08-915-096A-13 Sequence 13, Appl  
29 39 78.0 48 4 US-09-101-544-15 Sequence 15, Appl  
30 39 78.0 48 6 5434135-3 Patent No. 5434135  
31 39 78.0 51 6 5177197-50 Patent No. 5177197  
32 39 78.0 53 1 US-07-869-176-1 Sequence 1, Appl  
33 39 78.0 53 1 US-08-284-923-1 Sequence 1, Appl  
34 39 78.0 53 1 US-08-360-841-2 Sequence 2, Appl  
35 39 78.0 53 2 US-08-861-000-1 Sequence 1, Appl  
36 39 78.0 53 4 US-08-619-032B-1 Sequence 1, Appl  
37 39 78.0 53 6 5332669-3 Patent No. 5332669  
38 39 78.0 53 6 5434135-2 Patent No. 5434135  
39 39 78.0 55 6 5218093-1 Patent No. 5218093  
40 39 78.0 88 2 US-07-885-089B-13 Sequence 13, Appl  
41 39 78.0 91 1 US-07-847-743B-15 Sequence 15, Appl  
42 39 78.0 91 1 US-08-456-201-15 Sequence 15, Appl  
43 39 78.0 91 2 US-08-330-161-13 Sequence 13, Appl  
44 39 78.0 91 2 US-08-456-241-15 Sequence 15, Appl  
45 39 78.0 91 2 US-08-440-401-13 Sequence 13, Appl

## ALIGNMENTS

RESULT 1  
US-08-597-545-7  
; Sequence 7, Application US/08597545  
; Patent No. 5580738  
; GENERAL INFORMATION:  
; APPLICANT: LABORDA, Jorge  
; TITLE OF INVENTION: Delta-Like Gene Expressed In  
; TITLE OF INVENTION: Neuroendocrine Tumors  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/597,545  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/989,537  
; FILING DATE: 11-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 40399/166 NIHD  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-597-545-7

Query Match 100.0%; Score 50; DB 1; Length 12;  
Best Local Similarity 80.0%; Pred. No. 0.11;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

Db 3 CVIGSGDRC 12

RESULT 2  
US-08-457-135-7  
; Sequence 7, Application US/08457135  
; Patent No. 5644031  
; GENERAL INFORMATION:  
; APPLICANT: LABORDA, Jorge  
; TITLE OF INVENTION: Delta-Like Gene Expressed In  
; TITLE OF INVENTION: Neuroendocrine Tumors  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/457,135  
; FILING DATE: 01-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/989,537  
; FILING DATE: 11-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 40399/304/NIHD  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-457-135-7

Query Match 100.0%; Score 50; DB 1; Length 12;  
Best Local Similarity 80.0%; Pred. No. 0.11;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
| | | | |  
Db 3 CVIGYSGDRC 12

RESULT 3  
US-08-284-923-2  
; Sequence 2, Application US/08284923  
; Patent No. 5547935  
; GENERAL INFORMATION:  
; APPLICANT: Mullenbach, Guy T  
; APPLICANT: Blaney, Jeffrey M  
; APPLICANT: Rosenberg, Steven  
; TITLE OF INVENTION: Muteins of Epidermal Growth Factor  
; TITLE OF INVENTION: exhibiting enhanced binding at low ph  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/284,923  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/746,651  
; FILING DATE: 16-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mcclung, Barbara G  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 231,001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-601-2708  
; TELEFAX: 510-655-3542  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 53 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-284-923-2

Query Match 100.0%; Score 50; DB 1; Length 53;  
Best Local Similarity 80.0%; Pred. No. 0.44;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
| | | | |  
Db 33 CVIGYSGDRC 42

RESULT 4  
US-08-619-032B-2  
; Sequence 2, Application US/08619032B  
; Patent No. 6191106  
; GENERAL INFORMATION:  
; APPLICANT: Mullenbach, Guy T.  
; APPLICANT: Blaney, Jeffrey M.  
; APPLICANT: Rosenberg, Steven  
; TITLE OF INVENTION: MUTEINS OF EPIDERMAL GROWTH FACTOR  
; TITLE OF INVENTION: EXHIBITING ENHANCED BINDING AT LOW PH  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: Intellectual Property R-440, P.O. Box 8097  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94662-8097  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/619,032B  
; FILING DATE: 20-MAR-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Guth, Joseph H.  
; REGISTRATION NUMBER: 31,261  
; REFERENCE/DOCKET NUMBER: 0231,004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 923-3888  
; TELEFAX: (510) 655-3542  
; TELEX: N/A  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 53 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-619-032B-2

Query Match      100.0%; Score 50; DB 4; Length 53;
Best Local Similarity 80.0%; Pred. No. 0.44;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 CVIGXSGDXC 10
      ||||:||||:|
Db      33 CVIGYSGDRC 42

RESULT 5
5332669-2
;PATENT NO. 5332669
; APPLICANT: DEUEL, THOMAS F.
; TITLE OF INVENTION: PROSTATE-DERIVED MITOGEN
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/356,739
; FILING DATE: 24-MAR-1989
; SEQ ID NO: 2:
; LENGTH: 53
5332669-2

Query Match      100.0%; Score 50; DB 6; Length 53;
Best Local Similarity 80.0%; Pred. No. 0.44;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 CVIGXSGDXC 10
      ||||:||||:|
Db      33 CVIGYSGDRC 42

RESULT 6
5180808-2
;PATENT NO. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.
; TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN
; ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441,179
; FILING DATE: 27-NOV-1989
; SEQ ID NO: 2:
; LENGTH: 2409
5180808-2

Query Match      86.0%; Score 43; DB 6; Length 2409;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVIGXSGDXC 10
      || |:|||:|
Db      2128 CVPGYSGDQC 2137

RESULT 7
US-08-185-432-16
; Sequence 16, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
; ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:

;
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741

;
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
;
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-16

Query Match      86.0%; Score 43; DB 1; Length 2471;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVIGXSGDXC 10
      || |:|||:|
Db      326 CVNGWGGDC 335

RESULT 8
US-08-083-590A-19
; Sequence 19, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
```

TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2471 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-083-590A-19

Query Match 86.0%; Score 43; DB 1; Length 2471;  
Best Local Similarity 70.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|||:||||:  
Db 326 CVNGWSGDDC 335

## RESULT 9

US-08-532-384-19  
Sequence 19, Application US/08532384  
Patent No. 6083904  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And  
TITLE OF INVENTION: Nucleic Acids  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
CITY: 1155 Avenue of the Americas  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/532,384  
FILING DATE:  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/083,590  
FILING DATE: 25-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2471 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-532-384-19

Query Match 86.0%; Score 43; DB 3; Length 2471;  
Best Local Similarity 70.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|||:||||:  
Db 326 CVNGWSGDDC 335

RESULT 10  
US-08-899-232-1  
Sequence 1, Application US/08899232  
Patent No. 6436650  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Qi, Huilin  
TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON  
FILE REFERENCE: 7326-046  
CURRENT APPLICATION NUMBER: US/08/899,232  
CURRENT FILING DATE: 1997-07-23  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 2471  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-08-899-232-1

Query Match 86.0%; Score 43; DB 4; Length 2471;  
Best Local Similarity 70.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|||:||||:  
Db 326 CVNGWSGDDC 335

## RESULT 11

US-09-191-647-10  
Sequence 10, Application US/09191647  
Patent No. 6046015  
GENERAL INFORMATION:  
APPLICANT: Goodman, Corey  
APPLICANT: Kid, Thomas  
APPLICANT: Brose, Katja  
APPLICANT: Tessier-Lavigne, Marc  
TITLE OF INVENTION: Modulating Robo: Ligand Interactions  
FILE REFERENCE: B98-031-3  
CURRENT APPLICATION NUMBER: US/09/191,647  
CURRENT FILING DATE: 1998-11-13  
EARLIER APPLICATION NUMBER: 60/065,544  
EARLIER FILING DATE: 1997-11-14  
EARLIER APPLICATION NUMBER: 60/081,057  
EARLIER FILING DATE: 1998-04-07  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 10  
LENGTH: 154  
TYPE: PRT  
ORGANISM: mouse  
US-09-191-647-10

Query Match 84.0%; Score 42; DB 3; Length 154;  
Best Local Similarity 50.0%; Pred. No. 18;  
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|||:||||:  
Db 11 CMLGYTGDMC 20

## RESULT 12

US-09-540-245A-10  
Sequence 10, Application US/09540245A  
Patent No. 6270984  
GENERAL INFORMATION:  
APPLICANT: Goodman, Corey  
APPLICANT: Kid, Thomas  
APPLICANT: Brose, Katja  
APPLICANT: Tessier-Lavigne, Marc  
TITLE OF INVENTION: Modulating Robo: Ligand Interactions

QY 1 CVIGXSGDXC 10  
|||:||||:  
Db 326 CVNGWSGDDC 335

Query Match 80.0%; Score 40; DB 6; Length 48;  
Best Local Similarity 60.0%; Pred. No. 12;

Search completed: July 2, 2003, 07:25:44  
Job time : 12 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:28:49 ; Search time 103 Seconds  
(without alignments)  
11.164 Million cell updates/sec

Title: US-09-673-785D-13  
Perfect score: 50  
Sequence: 1 CVIGSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues  
Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US03\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	50	100.0	53	9 US-10-138-158-19	Sequence 19, Appl
2	50	100.0	53	9 US-10-150-648B-33	Sequence 33, Appl
3	50	100.0	145	9 US-10-150-648B-30	Sequence 30, Appl
4	43	86.0	176	9 US-10-024-599-4	Sequence 4, Appl
5	43	86.0	741	10 US-09-925-301-930	Sequence 930, App
6	42	84.0	572	9 US-09-900-449A-7	Sequence 7, Appl
7	42	84.0	601	9 US-09-900-449A-5	Sequence 5, Appl
8	42	84.0	639	9 US-09-900-449A-4	Sequence 4, Appl
9	40	80.0	112	9 US-10-211-994-20	Sequence 20, Appl
10	40	80.0	298	10 US-09-864-675-4	Sequence 4, Appl
11	40	80.0	878	9 US-10-157-031-171	Sequence 171, App
12	39	78.0	46	9 US-10-136-573A-12	Sequence 12, Appl
13	39	78.0	46	9 US-09-877-665-12	Sequence 12, Appl
14	39	78.0	46	9 US-10-215-862-12	Sequence 12, Appl
15	39	78.0	46	10 US-09-817-647-12	Sequence 12, Appl
16	39	78.0	47	12 US-10-096-241-17	Sequence 17, Appl
17	39	78.0	48	9 US-10-201-945-13	Sequence 13, Appl
18	39	78.0	53	9 US-09-903-327A-9	Sequence 9, Appl
19	39	78.0	53	9 US-10-150-648B-35	Sequence 35, Appl

Sequence 4, Appl  
Sequence 31, Appl  
Sequence 3, Appl  
Sequence 13, Appl  
Sequence 27, Appl  
Sequence 29, Appl  
Sequence 5, Appl  
Sequence 7, Appl  
Sequence 25, Appl  
Sequence 64, Appl  
Sequence 32, Appl  
Sequence 692, App  
Sequence 15, Appl  
Sequence 15, Appl  
Sequence 5, Appl  
Sequence 13, Appl  
Sequence 6, Appl  
Sequence 6, Appl  
Sequence 114, App  
Sequence 1126, Ap  
Sequence 6253, Ap  
Sequence 3, Appl  
Sequence 2, Appl  
Sequence 29, Appl  
Sequence 269, App

ALIGNMENTS

RESULT 1  
US-10-138-158-19  
; Sequence 19, Application US/10138158  
; Publication No. US20030036509A1  
; GENERAL INFORMATION:  
; APPLICANT: STEM CELL PHARMACEUTICALS, INC.  
; APPLICANT: TWARDZIK, Daniel R.  
; APPLICANT: PERNET, Andre  
; APPLICANT: FELKER, Thomas S.  
; APPLICANT: PASKELL, Stefan  
; APPLICANT: RENO, John M.  
; TITLE OF INVENTION: TGF-alpha POLYPEPTIDES, FUNCTIONAL FRAGMENTS AND METHODS OF  
; FILE REFERENCE: STEM1110-6  
; CURRENT APPLICATION NUMBER: US/10/138,158  
; CURRENT FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: US 09/641,587  
; PRIOR FILING DATE: 2000-08-17  
; PRIOR APPLICATION NUMBER: US 09/559,248  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: US 09/459,813  
; PRIOR FILING DATE: 1999-12-13  
; PRIOR APPLICATION NUMBER: US 09/492,935  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 09/378,567  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19  
; LENGTH: 53  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-138-158-19

Query Match 100.0%; Score 50; DB 9; Length 53;  
Best Local Similarity 80.0%; Pred. No. 0.51;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CVIGSGDXC 10  
Db 33 CVIGSGDRC 42

```
RESULT 2
US-10-150-648B-33
; Sequence 33, Application US/10150648B
; Publication No. US20030059802A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau-Goesseels, Sylvie
; APPLICANT: John, Sushil J.
; APPLICANT: Selinger, Leonard B.
; APPLICANT: Benkel, Bernhard F.
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
; TITLE OF INVENTION: factor
; FILE REFERENCE: 60-01
; CURRENT APPLICATION NUMBER: US/10/150,648B
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/292,136
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a
; TITLE: 128,000-molecular weight protein precursor
; JOURNAL: Nature
; VOLUME: 303
; PAGES: 722-725
; DATE: 1983
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380
; DATABASE ENTRY DATE: 1993-04-27
; RELEVANT RESIDUES: Relevant residues FROM 977 TO 1029
US-10-150-648B-33

Query Match      100.0%; Score 50; DB 9; Length 53;
Best Local Similarity 80.0%; Pred. No. 0.51;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
   |||||:|:|:|
Db 33 CVIGYSGDRC 42

RESULT 3
US-10-150-648B-30
; Sequence 30, Application US/10150648B
; Publication No. US20030059802A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau-Goesseels, Sylvie
; APPLICANT: John, Sushil J.
; APPLICANT: Selinger, Leonard B.
; APPLICANT: Benkel, Bernhard F.
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
; TITLE OF INVENTION: factor
; FILE REFERENCE: 60-01
; CURRENT APPLICATION NUMBER: US/10/150,648B
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/292,136
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a
; TITLE: 128,000-molecular weight protein precursor
; JOURNAL: Nature
; VOLUME: 303
; PAGES: 722-725
```

```
; DATE: 1983
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380
; DATABASE ENTRY DATE: 1993-04-27
; RELEVANT RESIDUES: Relevant residues FROM 919 TO 1063
US-10-150-648B-30

Query Match      100.0%; Score 50; DB 9; Length 145;
Best Local Similarity 80.0%; Pred. No. 1.3;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
   |||||:|:|:|
Db 91 CVIGYSGDRC 100

RESULT 4
US-10-024-599-4
; Sequence 4, Application US/10024599
; Patent No. US2002016532A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Cimborra, Daniel M.
; APPLICANT: Heichman, Karen.
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 2318-278-II
; CURRENT APPLICATION NUMBER: US/10/024,599
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/256,986
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: 1..176
; OTHER INFORMATION: Xaa is Gly or Cys
US-10-024-599-4

Query Match      86.0%; Score 43; DB 9; Length 176;
Best Local Similarity 70.0%; Pred. No. 18;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
   |||:|:|:|
Db 11 CVNGWGGDDC 20

RESULT 5
US-09-925-301-930
; Sequence 930, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 930
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
```

; LOCATION: (8)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (19)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (282)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-301-930

Query Match 86.0%; Score 43; DB 10; Length 741;  
Best Local Similarity 70.0%; Pred. No. 67;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10  
|||:|:|:|  
Db 460 CVPYSGDQC 469

## RESULT 6

US-09-900-449A-7  
; Sequence 7, Application US/09900449A  
; Publication No. US20030040616A1

## ; GENERAL INFORMATION:

; APPLICANT: ZHONG, Jenny et al.  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND  
; TITLE OF INVENTION: USES THEREOF

; FILE REFERENCE: CL001271  
; CURRENT APPLICATION NUMBER: US/09/900,449A  
; CURRENT FILING DATE: 2001-07-09

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7

; LENGTH: 572

; TYPE: PRT

; ORGANISM: Strongylocentrotus purpuratus

US-09-900-449A-7

Query Match 84.0%; Score 42; DB 9; Length 572;  
Best Local Similarity 60.0%; Pred. No. 76;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10  
|||:|:|:|  
Db 141 CPLGFGSDNC 150

## RESULT 7

US-09-900-449A-5  
; Sequence 5, Application US/09900449A  
; Publication No. US20030040616A1

## ; GENERAL INFORMATION:

; APPLICANT: ZHONG, Jenny et al.  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND  
; TITLE OF INVENTION: USES THEREOF

; FILE REFERENCE: CL001271

; CURRENT APPLICATION NUMBER: US/09/900,449A

; CURRENT FILING DATE: 2001-07-09

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 601

; TYPE: PRT

; ORGANISM: Strongylocentrotus purpuratus

US-09-900-449A-5

Query Match 84.0%; Score 42; DB 9; Length 601;  
Best Local Similarity 60.0%; Pred. No. 79;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

Db 132 CPLGFGSDNC 141  
|||:|:|:|

## RESULT 8

US-09-900-449A-4  
; Sequence 4, Application US/09900449A  
; Publication No. US20030040616A1

## ; GENERAL INFORMATION:

; APPLICANT: ZHONG, Jenny et al.

; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND

; TITLE OF INVENTION: USES THEREOF

; FILE REFERENCE: CL001271

; CURRENT APPLICATION NUMBER: US/09/900,449A

; CURRENT FILING DATE: 2001-07-09

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 639

; TYPE: PRT

; ORGANISM: Strongylocentrotus purpuratus

US-09-900-449A-4

Query Match 84.0%; Score 42; DB 9; Length 639;  
Best Local Similarity 60.0%; Pred. No. 84;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10  
|||:|:|:|  
Db 60 CPLGFGSDNC 69

## RESULT 9

US-10-211-994-20  
; Sequence 20, Application US/10211994  
; Publication No. US20030082201A1

## ; GENERAL INFORMATION:

; APPLICANT: Rao, M.R.S.

; APPLICANT: Sengupta, Paromita

; APPLICANT: Prasad, Sudhanand

; APPLICANT: Burman, Anand C.

; APPLICANT: Mukherjee, Rama

; APPLICANT: Thomas, Becky

; TITLE OF INVENTION: MULTIVALENT SYNTHETIC VACCINE FOR CANCER

; FILE REFERENCE: U014152-1

; CURRENT APPLICATION NUMBER: US/10/211,994

; CURRENT FILING DATE: 2002-08-02

; PRIOR APPLICATION NUMBER: 60/309,975

; PRIOR FILING DATE: 2001-08-03

; NUMBER OF SEQ ID NOS: 29

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 20

; LENGTH: 112

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Nucleotide and peptide sequence of multivalent vaccine

US-10-211-994-20

Query Match 80.0%; Score 40; DB 9; Length 112;  
Best Local Similarity 60.0%; Pred. No. 35;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10  
|||:|:|:|  
Db 92 CVIGYIGERC 101

## RESULT 10

US-09-864-675-4  
; Sequence 4, Application US/09864675  
; Patent No. US20020081286A1

GENERAL INFORMATION:  
APPLICANT: Marchionni, Mark  
TITLE OF INVENTION: NRG-2 NUCLEIC ACID MOLECULES,  
FILE REFERENCE: 04585/049002 POLYPEPTIDES, AND DIAGNOSTIC AND THERAPEUTIC METHODS  
CURRENT APPLICATION NUMBER: US/09/864,675  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/206,495  
PRIOR FILING DATE: 2000-05-23  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 298  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-864-675-4

Query Match 80.0%; Score 40; DB 10; Length 298;  
Best Local Similarity 50.0%; Pred. No. 85;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|:|:|:|:|  
Db 280 CPVGYTGDCR 289

## RESULT 11

US-10-157-031-171  
Sequence 171, Application US/10157031  
Publication No. US20030108890A1  
GENERAL INFORMATION:

APPLICANT: Baranova, A. V.  
APPLICANT: Yankovsky, N. K.  
APPLICANT: Kozlov, A. P.  
APPLICANT: Lobashev, A. V.  
APPLICANT: Krukovskaya, L. L.

TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences  
FILE REFERENCE: 2760-103  
CURRENT APPLICATION NUMBER: US/10/157,031  
CURRENT FILING DATE: 2002-05-30  
NUMBER OF SEQ ID NOS: 415  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 171  
LENGTH: 878  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-157-031-171

Query Match 80.0%; Score 40; DB 9; Length 878;  
Best Local Similarity 60.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|:|:|:|:|  
Db 526 CLPGFSGDRC 535

## RESULT 12

US-10-136-573A-12  
Sequence 12, Application US/10136573A  
Patent No. US20020161200A1  
GENERAL INFORMATION:

APPLICANT: Godowski, Paul J.  
APPLICANT: Zhang, Dong Xiao

TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and  
FILE REFERENCE: P1084R1C2  
CURRENT APPLICATION NUMBER: US/10/136,573A  
CURRENT FILING DATE: 2002-04-29  
PRIOR APPLICATION NUMBER: US 09/480,977  
PRIOR FILING DATE: 2000-01-11  
PRIOR APPLICATION NUMBER: US 08/899,437

PRIOR FILING DATE: 1997-07-24  
PRIOR APPLICATION NUMBER: US 60/052,019  
PRIOR FILING DATE: 1997-07-09  
NUMBER OF SEQ ID NOS: 23  
SEQ ID NO 12  
LENGTH: 46  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-136-573A-12

Query Match 78.0%; Score 39; DB 9; Length 46;  
Best Local Similarity 50.0%; Pred. No. 22;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|:|:|:|:|  
Db 32 CVVGYIGRC 41

## RESULT 13

US-09-877-665-12  
Sequence 12, Application US/09877665  
Patent No. US20020164680A1  
GENERAL INFORMATION:

APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao  
TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related  
Ligands and Uses Therefor

NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA: US/09/877,665  
APPLICATION NUMBER: US/09/877,665  
FILING DATE: 08-Jun-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/109,206  
FILING DATE: 30-Jun-1998

ATTORNEY/AGENT INFORMATION:  
NAME: Conley, Deirdre L.  
REGISTRATION NUMBER: 36,487  
REFERENCE/DOCKET NUMBER: P1084R1-1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-2066  
TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear

FEATURE:  
NAME/KEY: hEGF.esf  
LOCATION: 1-46  
IDENTIFICATION METHOD:  
OTHER INFORMATION:

SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-09-877-665-12

Query Match 78.0%; Score 39; DB 9; Length 46;  
Best Local Similarity 50.0%; Pred. No. 22;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|:|:|:|:|

Db 32 CVVGYIGERC 41

## RESULT 14

US-10-215-862-12

; Sequence 12, Application US/10215862

; Publication No. US2003036166A1

; GENERAL INFORMATION:

; APPLICANT: Godowski, Paul J.

; APPLICANT: Mark, Melanie Rose

; APPLICANT: Zhang, Dong Xiao

; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and

; FILE OF INVENTION: Uses Therefor

; FILE REFERENCE: P1084RID2C1

; CURRENT APPLICATION NUMBER: US/10/215,862

; CURRENT FILING DATE: 2002-09-24

; PRIOR APPLICATION NUMBER: US 09/126,663

; PRIOR FILING DATE: 1998-07-30

; PRIOR APPLICATION NUMBER: US 08/899,437

; PRIOR FILING DATE: 1997-07-24

; PRIOR APPLICATION NUMBER: US 60/052,019

; PRIOR FILING DATE: 1997-07-09

; NUMBER OF SEQ ID NOS: 23

; SEQ ID NO 12

; LENGTH: 46

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-215-862-12

Query Match

78.0%; Score 39; DB 9; Length 46;

Best Local Similarity 50.0%; Pred. No. 22;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

||:|: |::|

Db 32 CVVGYIGERC 41

## RESULT 15

US-09-817-647-12

; Sequence 12, Application US/09817647

; Patent No. US20020082229A1

; GENERAL INFORMATION:

; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao

; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 1 DNA Way

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WinPatin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/817,647

; FILING DATE: 26-Mar-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/107,979

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Conley, Deirdre L.

; REGISTRATION NUMBER: 36,487

; REFERENCE/DOCKET NUMBER: P1084R1-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650/225-2066

; TELEFAX: 650/952-9881

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 46 amino acids

; TYPE: Amino Acid

; TOPOLOGY: Linear

; FEATURE:

; NAME/KEY: heGF.egf

; LOCATION: 1-46

; IDENTIFICATION METHOD:

; OTHER INFORMATION:

; SEQUENCE DESCRIPTION: SEQ ID NO: 12:

US-09-817-647-12

Query Match

78.0%; Score 39; DB 10; Length 46;

Best Local Similarity 50.0%; Pred. No. 22;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

||:|: |::|

Db 32 CVVGYIGERC 41

Search completed: July 2, 2003, 07:57:19

Job time : 104 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 12 seconds  
(without alignments)  
80.112 Million cell updates/sec

Title: US-09-673-785D-13  
Perfect score: 50  
Sequence: 1 CVIGXSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	1217	1 EGMSMG	epidermal growth f
2	45	90.0	2139	2 A35672	crumbs protein - f
3	45	90.0	4006	2 T09070	probable tenascin
4	43	86.0	102	2 B55885	chondroitin sulfat
5	43	86.0	372	2 S29359	hypothetical prote
6	43	86.0	862	2 S43922	versican - pig-tai
7	43	86.0	1203	2 A49175	Notch B protein -
8	43	86.0	1643	2 T14274	versican precursor
9	43	86.0	2409	1 A60979	versican precursor
10	43	86.0	2471	2 A49128	cell-fate determin
11	43	86.0	2531	2 T31070	notch homolog - se
12	43	86.0	3381	2 T42389	versican precursor
13	42	84.0	1064	2 A40136	fibropellin Ia - s
14	41	82.0	2352	2 T30201	Notch homolog prot
15	40	80.0	57	2 FC4415	Erbb kinase activa
16	40	80.0	161	2 S03938	neuroendocrine pro
17	40	80.0	601	2 T48539	probable potassium
18	40	80.0	1133	1 EGRT	epidermal growth f
19	40	80.0	1531	2 T42218	slit-1 protein hom
20	40	80.0	2437	2 S42612	transmembrane prot
21	39	78.0	264	2 T16271	hypothetical prote
22	39	78.0	379	2 T16213	APX-1 protein homo
23	39	78.0	422	2 S47581	hypothetical prote
24	39	78.0	426	2 AH1831	hypothetical prote
25	39	78.0	470	2 F85164	hypothetical prote
26	39	78.0	601	2 T34396	hypothetical prote
27	39	78.0	832	2 A31246	neurogenic protein
28	39	78.0	833	2 S19087	gene Delta protein
29	39	78.0	880	2 S00670	neurogenic repetit

#### ALIGNMENTS

##### RESULT 1

##### EGMSMG

epidermal growth factor precursor - mouse

N:Alternate names: urogastrone precursor

C:Species: Mus musculus (house mouse)

C:Date: 30-Nov-1980 #sequence-revision 11-Aug-1983 #text\_change 19-Jan-2001

C:Accession: A94272; A93304; A92118; A01387

R:Scott, J.; Urdeda, M.; Quiroga, M.; Sanchez-Pescador, R.; Fong, N.; Selby, M.; Rut

Science 221, 236-240, 1983

A:Title: Structure of a mouse submaxillary messenger RNA encoding epidermal growth

A:Reference number: A94272; MUID:83223630; PMID:6602382

A:Accession: A94272

A:Molecule type: mRNA

A:Residues: 1-1217 <SCO>

A:Cross-references: GB:J00380; NID:g192993; PIDN:AAA37539.1; PID:g309210

R:Gray, A.; Dull, T.; Ullrich, A.

Nature 303, 722-725, 1983

A:Title: Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-mol

A:Reference number: A93304; MUID:83219309; PMID:6304537

A:Accession: A93304

A:Molecule type: protein

A:Residues: 1-789,'Y',791-1047,'S',1049-1168 <GRA>

A:Cross-references: GB:J00380

A:Note: the sequence shown by these authors differs from residues 1134-1168 due to

ence of Scott et al.)

R:Savage Jr., C.R.; Inagami, T.; Cohen, S.

J. Biol. Chem. 247, 7612-7621, 1972

A:Title: The primary structure of epidermal growth factor.

A:Reference number: A92118; MUID:73048516; PMID:4636327

A:Accession: A92118

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Contents: annotation; disulfide bonds

C:Comment: Epidermal growth factor (EGF) stimulates the proliferation and different

gastrointestinal cell proliferation.

C:Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound p

C:Comment: The active growth factor from this submaxillary gland protein stimulates

C:Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-cr

C:Keywords: duplication; growth factor; tandem repeat; transmembrane protein

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-1217/Product: epidermal growth factor proprotein, membrane-bound form #status

F:29-1038/Domain: extracellular #status predicted <EXT>

F:50-485/Region: EGF precursor long repeat

F:53-92/Domain: LDL receptor YWTD-containing repeat homology <YW01>

F:93-134/Domain: LDL receptor YWTD-containing repeat homology <YW02>

brevican precursor  
ATP-dependent perm  
epidermal growth f  
protein FilC7.4 [i  
versican precursor  
hypothetical prote  
zonadhesin - pig  
tenascin-X precurs  
tenascin-X - bovin  
hypothetical prote  
bactracin synthet  
hydrophobin Ceg-2  
hypothetical prote  
conserved hypothet  
hypothetical prote  
mucin - rat (fragm

30 39 78.0 883 2 S49126  
31 39 78.0 1049 1 S19421  
32 39 78.0 1207 1 EGHU  
33 39 78.0 1722 2 E9753  
34 39 78.0 2397 1 A55535  
35 39 78.0 2824 2 T22759  
36 38 76.0 2476 2 T34022  
37 38 76.0 3566 1 A40701  
38 38 76.0 4135 2 T42629  
39 38 76.0 4307 2 T20721  
40 38 76.0 5255 2 T31677  
41 37 74.0 108 2 A46222  
42 37 74.0 294 2 E98236  
43 37 74.0 294 2 AF3049  
44 37 74.0 373 2 T22300  
45 37 74.0 447 2 A39321

F:135-176/Domain: LDL receptor WYTD-containing repeat homology <YW03>  
F:177-217/Domain: LDL receptor WYTD-containing repeat homology <YW04>  
F:218-262/Domain: LDL receptor WYTD-containing repeat homology <YW05>  
F:263-307/Domain: LDL receptor WYTD-containing repeat homology <YW06>  
F:324-360/Domain: EGF homology #status atypical <EG1>  
F:366-401/Domain: EGF homology <EG2>  
F:407-442/Domain: EGF homology <EG3>  
F:445-482/Domain: EGF homology <EG4>  
F:486-961/Region: EGF precursor long repeat  
F:489-529/Domain: LDL receptor WYTD-containing repeat homology <YW07>  
F:530-572/Domain: LDL receptor WYTD-containing repeat homology <YW08>  
F:573-615/Domain: LDL receptor WYTD-containing repeat homology <YW09>  
F:616-659/Domain: LDL receptor WYTD-containing repeat homology <YW10>  
F:660-700/Domain: LDL receptor WYTD-containing repeat homology <YW11>  
F:701-743/Domain: LDL receptor WYTD-containing repeat homology <YW12>  
F:751-786/Domain: EGF homology <EG5>  
F:842-875/Domain: EGF homology <EG6>  
F:881-917/Domain: EGF homology <EG7>  
F:923-958/Domain: EGF homology <EG8>  
F:977-1029/Product: epidermal growth factor #status experimental <EGF>  
F:982-1018/Domain: EGF homology <EG9>  
F:1039-1063/Domain: transmembrane #status predicted <TM>  
F:1064-1217/Domain: intracellular #status predicted <INT>  
F:347-360,366-377,373-386,388-401,407-418,414-427,429-442,445-457,453-467,469-482,751-76  
tus predicted  
F:982-996,990-1007,1009-1018/Disulfide bonds: #status experimental

Query Match 100.0%; Score 50; DB 1; Length 1217;  
Best Local Similarity 80.0%; Pred. No. 1.6;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|:|:|:|:|:  
DB 1009 CVIGSGDRC 1018

## RESULT 2

A35672  
crumbs protein - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 21-Sep-1990 #sequence\_revision 18-Nov-1992 #text\_change 11-Jan-2000  
C:Accession: A35672  
Cell 61, 787-799, 1990  
R:Teppas, U.; Theres, C.; Knust, E.  
A:Title: crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila  
A:Reference number: A35672; MUID:90263104; PMID:2344615  
A:Accession: A35672  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2139 <TEP>  
A:Cross-references: GB:M33753  
A:Note: the authors translated the codon GGC for residue 1928 as Cys, and TAT for residue  
C:Genetics:  
A:Gene: FlyBase:crb  
A:Cross-references: FlyBase:FBgn0000368  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
C:Keywords: transmembrane protein  
F:352-385/Domain: EGF homology <EGX1>  
F:392-424/Domain: EGF homology <EGF1>  
F:691-722/Domain: EGF homology <EGF>  
F:767-799/Domain: EGF homology <EGF3>  
F:1878-1914/Domain: EGF homology <EGX2>

Query Match 90.0%; Score 45; DB 2; Length 2139;  
Best Local Similarity 60.0%; Pred. No. 20;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|:|:|:|:|:  
DB 571 CAVGSGDRC 580

## RESULT 3

T09070  
probable tenascin X - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jan-2000  
C:Accession: T09070  
R:Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.;  
submitted to the EMBL Data Library, October 1997  
A:Description: Sequence of the mouse major histocompatibility locus class III region.  
A:Reference number: Z16543  
A:Accession: T09070  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-4006 <ROW>  
A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958  
C:Genetics:  
A:Gene: TNX  
A:Map position: 17  
A:introns: 124/1; 735/1; 773/3; 826/1; 914/1; 1037/1; 1135/1; 1232/1; 1329/1; 144C/1;  
019/1; 3115/1; 3208/1; 3302/1; 3405/1; 3517/1; 3558/1; 3606/1; 3646/1; 3694/1; 3757/3  
C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin  
C:Keywords: extracellular matrix  
F:423-448/Domain: EGF homology <EGF>  
F:826-906/Domain: fibronectin type III repeat homology <3FR>  
F:7389-3997/Domain: fibrinogen beta/gamma homology <FBG>

Query Match 90.0%; Score 45; DB 2; Length 4006;  
Best Local Similarity 60.0%; Pred. No. 37;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|:|:|:|:|:  
DB 532 CAVGSGDDC 541

## RESULT 4

B55885  
chondroitin sulfate proteoglycan MV3 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 16-Jul-1999  
C:Accession: B55885  
R:Zako, M.; Shinomura, T.; Ujita, M.; Ito, K.; Kimata, K.  
J. Biol. Chem. 270, 3914-3918, 1995  
A:Title: Expression of PG-M(V3), an alternatively spliced form of PG-M without a cho  
A:Reference number: A55885; MUID:95181355; PMID:7876137  
A:Accession: B55885  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-102 <ZAK>  
A:Cross-references: GB:S75879; GB:D32039  
C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology;  
F:24-55/Domain: EGF homology <EG1>  
F:62-93/Domain: EGF homology <EG2>

Query Match 86.0%; Score 43; DB 2; Length 102;  
Best Local Similarity 70.0%; Pred. No. 2.5;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
|:|:|:|:|:  
DB 46 CVPGYSGDQC 55

## RESULT 5

T29359  
hypothetical protein R05G6.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T29359  
R:Murray, J.; Le, T.T.  
submitted to the EMBL Data Library, May 1996  
A:Description: The sequence of C. elegans cosmid R05G6.  
A:Reference number: Z20612  
A:Accession: T29359



A;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-372 <MUR>

A:Cross-references: EMBL:U58746; PIDN:AAB00626.1; GSPDB:GN00022; CESP:R05G6.9

A:Experimental source: strain Bristol N2; clone R05G6

C:Genetics:

A:Gene: CESP:R05G6.9

A:Map position: 4

A:Introns: 80/1; 161/1; 245/1; 286/1

Query Match 86.0%; Score 43; DB 2; Length 372;

Best Local Similarity 50.0%; Pred. No. 8.4;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10

|||:||||

Db 66 CFLGWSGDCS 75

RESULT 6

S43922

N:Alternate names: chondroitin sulfate proteoglycan

C:Species: Macaca nemestrina (pig-tailed macaque)

C:Date: 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 16-Jul-1999

C:Accession: S43922

R:Yao, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.

Matrix Biol. 14, 213-225, 1994

A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by

A:Reference number: S43921; MUID:95005762; PMID:7921538

A:Accession: S43922

A:Molecule type: mRNA

A:Residues: 1-233;234-525;526-862 <YAO>

A:Cross-references: EMBL:S72413

A:Note: 507-Ser was also found

A:Note: the authors translated the codon GCC for residue 50 as Val, AAG for residue 422

669 as Asn

C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EG

C:Keywords: chondroitin sulfate proteoglycan; extracellular matrix

F:1-37/Domain: link protein repeat homology (fragment) <LNK1>

F:58-139/Domain: link protein repeat homology <LNK2>

F:722-753/Domain: EGF homology <EG1>

F:760-791/Domain: EGF homology <EG2>

Query Match

Best Local Similarity 86.0%; Score 43; DB 2; Length 862;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10

|||:||||

Db 744 CVPGYSGDQC 753

RESULT 7

A49175

Motch B protein - mouse (fragment)

N:Alternate names: Notch homolog

C:Species: Mus musculus (house mouse)

C:Date: 21-Jan-1994 #sequence\_revision 05-Jan-1996 #text\_change 20-Sep-1999

C:Accession: A49175; PHL570; S32113

R:Bardele, M.; Lendahl, U.

Exp. Cell Res. 204, 364-372, 1993

A:Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety of

A:Reference number: A49175; MUID:93178563; PMID:8440332

A:Accession: A49175

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1203 <LAR>

A:Cross-references: EMBL:X68279; MID:g287989; PIDN:CAA48340.1; PID:g287990

A:Experimental source: embryo

A:Note: sequence extracted from NCBI backbone (NCBIP:126158)

C:Comment: This protein has many EGF repeats and lin-12/Notch repeats.

C:Comment: This protein is one of the neurogenic proteins controlling the decision betwe

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology  
F:143-174/Domain: EGF homology <EGX1>  
F:482-513/Domain: EGF homology <EGF1>  
F:560-591/Domain: EGF homology <EGF>  
F:674-705/Domain: EGF homology <EGX2>  
F:712-743/Domain: EGF homology <EGF3>  
F:836-867/Domain: EGF homology <EGX3>

Query Match 86.0%; Score 43; DB 2; Length 1203;

Best Local Similarity 70.0%; Pred No. 26;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

|||:||||

Db 9 CVNGWSGDDC 18

RESULT 8

T14274

versican precursor, splice form V2 - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 05-May-2000

C:Accession: T14274

R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.

J. Biol. Chem. 273, 15758-15764, 1998

A:Title: Versican V2 is a major extracellular matrix component of the mature bovine

A:Reference number: Z17954; MUID:98288320; PMID:9624174

A:Accession: T14274

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1643 <SCH>

A:Cross-references: EMBL:AF060458; MID:g3253303; PID:g3253304; PIDN:AAC24360.1

A:Experimental source: brain

C:Keywords: glycoprotein

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-1643/Product: versican, splice form V2 #status predicted <MAT>

F:57,331,352,817,965,1017,1333,1616,1636/Binding site: carbohydrate (Asn) (covalent)

Query Match

Best Local Similarity 86.0%; Score 43; DB 2; Length 1643;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10

|||:||||

Db 1362 CVPGYSGDRC 1371

RESULT 9

A60979

versican precursor - human

N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteogly

N:Contains: glial hyaluronate-binding protein

C:Species: Homo sapiens (man)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Jan-2001

C:Accession: S06014; S43921; A60979; A30358; A29348; A45131; I54179

R:Zimmermann, D.R.; Ruoslahti, E.

EMBO J. 8, 2975-2981, 1989

A:Title: Multiple domains of the large fibroblast proteoglycan, versican.

A:Reference number: S06014; MUID:90059882; PMID:2583089

A:Accession: S06014

A:Molecule type: mRNA

A:Residues: 1-2409 <ZIM>

A:Cross-references: GB:X15998; MID:g37662; PIDN:CAA34128.1; PID:g37663

R:Yao, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.

Matrix Biol. 14, 213-225, 1994

A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cell

A:Reference number: S43921; MUID:95005762; PMID:7921538

A:Accession: S43921

A:Molecule type: mRNA

A:Residues: 208-440;1094-1385;1910-2246 <YAO>

R:Bignami, A.; Lane, W.S.; Andrews, D.; Dahl, D.

Brain Res. Bull. 22, 67-70, 1989

A:Title: Structural similarity of hyaluronate binding proteins in brain and cartilag

A;Reference number: A60979; MUID:89229983; PMID:2469524  
A;Accession: A60979  
A;Molecule type: protein  
A;Residues: 171-210;289-303 <BIG>  
R;Perides, G.; Lane, W.S.; Andrews, D.; Dahl, D.; Bignami, A.  
J. Biol. Chem. 264, 5981-5987, 1989  
A;Title: Isolation and partial characterization of a glial hyaluronate-binding protein.  
A;Reference number: A30358; MUID:89174663; PMID:2466833  
A;Accession: A30358  
A;Molecule type: protein  
A;Residues: 24-50;80-87, D', 89-119;128-155;167-218;229-259, 'TR', 261-268;277-283, 'G', 285-289, 'G', 291-294, 'G', 295-296, 'G', 297-298, 'G', 299-300, 'G', 301-302, 'G', 303-304, 'G', 305-306, 'G', 307-308, 'G', 309-310, 'G', 311-312, 'G', 313-314, 'G', 315-316, 'G', 317-318, 'G', 319-320, 'G', 321-322, 'G', 323-324, 'G', 325-326, 'G', 327-328, 'G', 329-330, 'G', 331-332, 'G', 333-334, 'G', 335-336, 'G', 337-338, 'G', 339-340, 'G', 341-342, 'G', 343-344, 'G', 345-346, 'G', 347-348, 'G', 349-350, 'G', 351-352, 'G', 353-354, 'G', 355-356, 'G', 357-358, 'G', 359-360, 'G', 361-362, 'G', 363-364, 'G', 365-366, 'G', 367-368, 'G', 369-370, 'G', 371-372, 'G', 373-374, 'G', 375-376, 'G', 377-378, 'G', 379-380, 'G', 381-382, 'G', 383-384, 'G', 385-386, 'G', 387-388, 'G', 389-390, 'G', 391-392, 'G', 393-394, 'G', 395-396, 'G', 397-398, 'G', 399-400, 'G', 401-402, 'G', 403-404, 'G', 405-406, 'G', 407-408, 'G', 409-410, 'G', 411-412, 'G', 413-414, 'G', 415-416, 'G', 417-418, 'G', 419-420, 'G', 421-422, 'G', 423-424, 'G', 425-426, 'G', 427-428, 'G', 429-430, 'G', 431-432, 'G', 433-434, 'G', 435-436, 'G', 437-438, 'G', 439-440, 'G', 441-442, 'G', 443-444, 'G', 445-446, 'G', 447-448, 'G', 449-450, 'G', 451-452, 'G', 453-454, 'G', 455-456, 'G', 457-458, 'G', 459-460, 'G', 461-462, 'G', 463-464, 'G', 465-466, 'G', 467-468, 'G', 469-470, 'G', 471-472, 'G', 473-474, 'G', 475-476, 'G', 477-478, 'G', 479-480, 'G', 481-482, 'G', 483-484, 'G', 485-486, 'G', 487-488, 'G', 489-490, 'G', 491-492, 'G', 493-494, 'G', 495-496, 'G', 497-498, 'G', 499-500, 'G', 501-502, 'G', 503-504, 'G', 505-506, 'G', 507-508, 'G', 509-510, 'G', 511-512, 'G', 513-514, 'G', 515-516, 'G', 517-518, 'G', 519-520, 'G', 521-522, 'G', 523-524, 'G', 525-526, 'G', 527-528, 'G', 529-530, 'G', 531-532, 'G', 533-534, 'G', 535-536, 'G', 537-538, 'G', 539-540, 'G', 541-542, 'G', 543-544, 'G', 545-546, 'G', 547-548, 'G', 549-550, 'G', 551-552, 'G', 553-554, 'G', 555-556, 'G', 557-558, 'G', 559-560, 'G', 561-562, 'G', 563-564, 'G', 565-566, 'G', 567-568, 'G', 569-570, 'G', 571-572, 'G', 573-574, 'G', 575-576, 'G', 577-578, 'G', 579-580, 'G', 581-582, 'G', 583-584, 'G', 585-586, 'G', 587-588, 'G', 589-590, 'G', 591-592, 'G', 593-594, 'G', 595-596, 'G', 597-598, 'G', 599-600, 'G', 601-602, 'G', 603-604, 'G', 605-606, 'G', 607-608, 'G', 609-610, 'G', 611-612, 'G', 613-614, 'G', 615-616, 'G', 617-618, 'G', 619-620, 'G', 621-622, 'G', 623-624, 'G', 625-626, 'G', 627-628, 'G', 629-630, 'G', 631-632, 'G', 633-634, 'G', 635-636, 'G', 637-638, 'G', 639-640, 'G', 641-642, 'G', 643-644, 'G', 645-646, 'G', 647-648, 'G', 649-650, 'G', 651-652, 'G', 653-654, 'G', 655-656, 'G', 657-658, 'G', 659-660, 'G', 661-662, 'G', 663-664, 'G', 665-666, 'G', 667-668, 'G', 669-670, 'G', 671-672, 'G', 673-674, 'G', 675-676, 'G', 677-678, 'G', 679-680, 'G', 681-682, 'G', 683-684, 'G', 685-686, 'G', 687-688, 'G', 689-690, 'G', 691-692, 'G', 693-694, 'G', 695-696, 'G', 697-698, 'G', 699-700, 'G', 701-702, 'G', 703-704, 'G', 705-706, 'G', 707-708, 'G', 709-710, 'G', 711-712, 'G', 713-714, 'G', 715-716, 'G', 717-718, 'G', 719-720, 'G', 721-722, 'G', 723-724, 'G', 725-726, 'G', 727-728, 'G', 729-730, 'G', 731-732, 'G', 733-734, 'G', 735-736, 'G', 737-738, 'G', 739-740, 'G', 741-742, 'G', 743-744, 'G', 745-746, 'G', 747-748, 'G', 749-750, 'G', 751-752, 'G', 753-754, 'G', 755-756, 'G', 757-758, 'G', 759-760, 'G', 761-762, 'G', 763-764, 'G', 765-766, 'G', 767-768, 'G', 769-770, 'G', 771-772, 'G', 773-774, 'G', 775-776, 'G', 777-778, 'G', 779-780, 'G', 781-782, 'G', 783-784, 'G', 785-786, 'G', 787-788, 'G', 789-790, 'G', 791-792, 'G', 793-794, 'G', 795-796, 'G', 797-798, 'G', 799-800, 'G', 801-802, 'G', 803-804, 'G', 805-806, 'G', 807-808, 'G', 809-810, 'G', 811-812, 'G', 813-814, 'G', 815-816, 'G', 817-818, 'G', 819-820, 'G', 821-822, 'G', 823-824, 'G', 825-826, 'G', 827-828, 'G', 829-830, 'G', 831-832, 'G', 833-834, 'G', 835-836, 'G', 837-838, 'G', 839-840, 'G', 841-842, 'G', 843-844, 'G', 845-846, 'G', 847-848, 'G', 849-850, 'G', 851-852, 'G', 853-854, 'G', 855-856, 'G', 857-858, 'G', 859-860, 'G', 861-862, 'G', 863-864, 'G', 865-866, 'G', 867-868, 'G', 869-870, 'G', 871-872, 'G', 873-874, 'G', 875-876, 'G', 877-878, 'G', 879-880, 'G', 881-882, 'G', 883-884, 'G', 885-886, 'G', 887-888, 'G', 889-890, 'G', 891-892, 'G', 893-894, 'G', 895-896, 'G', 897-898, 'G', 899-900, 'G', 901-902, 'G', 903-904, 'G', 905-906, 'G', 907-908, 'G', 909-910, 'G', 911-912, 'G', 913-914, 'G', 915-916, 'G', 917-918, 'G', 919-920, 'G', 921-922, 'G', 923-924, 'G', 925-926, 'G', 927-928, 'G', 929-930, 'G', 931-932, 'G', 933-934, 'G', 935-936, 'G', 937-938, 'G', 939-940, 'G', 941-942, 'G', 943-944, 'G', 945-946, 'G', 947-948, 'G', 949-950, 'G', 951-952, 'G', 953-954, 'G', 955-956, 'G', 957-958, 'G', 959-960, 'G', 961-962, 'G', 963-964, 'G', 965-966, 'G', 967-968, 'G', 969-970, 'G', 971-972, 'G', 973-974, 'G', 975-976, 'G', 977-978, 'G', 979-980, 'G', 981-982, 'G', 983-984, 'G', 985-986, 'G', 987-988, 'G', 989-990, 'G', 991-992, 'G', 993-994, 'G', 995-996, 'G', 997-998, 'G', 999-1000, 'G', 1001-1002, 'G', 1003-1004, 'G', 1005-1006, 'G', 1007-1008, 'G', 1009-1010, 'G', 1011-1012, 'G', 1013-1014, 'G', 1015-1016, 'G', 1017-1018, 'G', 1019-1020, 'G', 1021-1022, 'G', 1023-1024, 'G', 1025-1026, 'G', 1027-1028, 'G', 1029-1030, 'G', 1031-1032, 'G', 1033-1034, 'G', 1035-1036, 'G', 1037-1038, 'G', 1039-1040, 'G', 1041-1042, 'G', 1043-1044, 'G', 1045-1046, 'G', 1047-1048, 'G', 1049-1050, 'G', 1051-1052, 'G', 1053-1054, 'G', 1055-1056, 'G', 1057-1058, 'G', 1059-1060, 'G', 1061-1062, 'G', 1063-1064, 'G', 1065-1066, 'G', 1067-1068, 'G', 1069-1070, 'G', 1071-1072, 'G', 1073-1074, 'G', 1075-1076, 'G', 1077-1078, 'G', 1079-1080, 'G', 1081-1082, 'G', 1083-1084, 'G', 1085-1086, 'G', 1087-1088, 'G', 1089-1090, 'G', 1091-1092, 'G', 1093-1094, 'G', 1095-1096, 'G', 1097-1098, 'G', 1099-1100, 'G', 1101-1102, 'G', 1103-1104, 'G', 1105-1106, 'G', 1107-1108, 'G', 1109-1110, 'G', 1111-1112, 'G', 1113-1114, 'G', 1115-1116, 'G', 1117-1118, 'G', 1119-1120, 'G', 1121-1122, 'G', 1123-1124, 'G', 1125-1126, 'G', 1127-1128, 'G', 1129-1130, 'G', 1131-1132, 'G', 1133-1134, 'G', 1135-1136, 'G', 1137-1138, 'G', 1139-1140, 'G', 1141-1142, 'G', 1143-1144, 'G', 1145-1146, 'G', 1147-1148, 'G', 1149-1150, 'G', 1151-1152, 'G', 1153-1154, 'G', 1155-1156, 'G', 1157-1158, 'G', 1159-1160, 'G', 1161-1162, 'G', 1163-1164, 'G', 1165-1166, 'G', 1167-1168, 'G', 1169-1170, 'G', 1171-1172, 'G', 1173-1174, 'G', 1175-1176, 'G', 1177-1178, 'G', 1179-1180, 'G', 1181-1182, 'G', 1183-1184, 'G', 1185-1186, 'G', 1187-1188, 'G', 1189-1190, 'G', 1191-1192, 'G', 1193-1194, 'G', 1195-1196, 'G', 1197-1198, 'G', 1199-1200, 'G', 1201-1202, 'G', 1203-1204, 'G', 1205-1206, 'G', 1207-1208, 'G', 1209-1210, 'G', 1211-1212, 'G', 1213-1214, 'G', 1215-1216, 'G', 1217-1218, 'G', 1219-1220, 'G', 1221-1222, 'G', 1223-1224, 'G', 1225-1226, 'G', 1227-1228, 'G', 1229-1230, 'G', 1231-1232, 'G', 1233-1234, 'G', 1235-1236, 'G', 1237-1238, 'G', 1239-1240, 'G', 1241-1242, 'G', 1243-1244, 'G', 1245-1246, 'G', 1247-1248, 'G', 1249-1250, 'G', 1251-1252, 'G', 1253-1254, 'G', 1255-1256, 'G', 1257-1258, 'G', 1259-1260, 'G', 1261-1262, 'G', 1263-1264, 'G', 1265-1266, 'G', 1267-1268, 'G', 1269-1270, 'G', 1271-1272, 'G', 1273-1274, 'G', 1275-1276, 'G', 1277-1278, 'G', 1279-1280, 'G', 1281-1282, 'G', 1283-1284, 'G', 1285-1286, 'G', 1287-1288, 'G', 1289-1290, 'G', 1291-1292, 'G', 1293-1294, 'G', 1295-1296, 'G', 1297-1298, 'G', 1299-1300, 'G', 1301-1302, 'G', 1303-1304, 'G', 1305-1306, 'G', 1307-1308, 'G', 1309-1310, 'G', 1311-1312, 'G', 1313-1314, 'G', 1315-1316, 'G', 1317-1318, 'G', 1319-1320, 'G', 1321-1322, 'G', 1323-1324, 'G', 1325-1326, 'G', 1327-1328, 'G', 1329-1330, 'G', 1331-1332, 'G', 1333-1334, 'G', 1335-1336, 'G', 1337-1338, 'G', 1339-1340, 'G', 1341-1342, 'G', 1343-1344, 'G', 1345-1346, 'G', 1347-1348, 'G', 1349-1350, 'G', 1351-1352, 'G', 1353-1354, 'G', 1355-1356, 'G', 1357-1358, 'G', 1359-1360, 'G', 1361-1362, 'G', 1363-1364, 'G', 1365-1366, 'G', 1367-1368, 'G', 1369-1370, 'G', 1371-1372, 'G', 1373-1374, 'G', 1375-1376, 'G', 1377-1378, 'G', 1379-1380, 'G', 1381-1382, 'G', 1383-1384, 'G', 1385-1386, 'G', 1387-1388, 'G', 1389-1390, 'G', 1391-1392, 'G', 1393-1394, 'G', 1395-1396, 'G', 1397-1398, 'G', 1399-1400, 'G', 1401-1402, 'G', 1403-1404, 'G', 1405-1406, 'G', 1407-1408, 'G', 1409-1410, 'G', 1411-1412, 'G', 1413-1414, 'G', 1415-1416, 'G', 1417-1418, 'G', 1419-1420, 'G', 1421-1422, 'G', 1423-1424, 'G', 1425-1426, 'G', 1427-1428, 'G', 1429-1430, 'G', 1431-1432, 'G', 1433-1434, 'G', 1435-1436, 'G', 1437-1438, 'G', 1439-1440, 'G', 1441-1442, 'G', 1443-1444, 'G', 1445-1446, 'G', 1447-1448, 'G', 1449-1450, 'G', 1451-1452, 'G', 1453-1454, 'G', 1455-1456, 'G', 1457-1458, 'G', 1459-1460, 'G', 1461-1462, 'G', 1463-1464, 'G', 1465-1466, 'G', 1467-1468, 'G', 1469-1470, 'G', 1471-1472, 'G', 1473-1474, 'G', 1475-1476, 'G', 1477-1478, 'G', 1479-1480, 'G', 1481-1482, 'G', 1483-1484, 'G', 1485-1486, 'G', 1487-1488, 'G', 1489-1490, 'G', 1491-1492, 'G', 1493-1494, 'G', 1495-1496, 'G', 1497-1498, 'G', 1499-1500, 'G', 1501-1502, 'G', 1503-1504, 'G', 1505-1506, 'G', 1507-1508, 'G', 1509-1510, 'G', 1511-1512, 'G', 1513-1514, 'G', 1515-1516, 'G', 1517-1518, 'G', 1519-1520, 'G', 1521-1522, 'G', 1523-1524, 'G', 1525-1526, 'G', 1527-1528, 'G', 1529-1530, 'G', 1531-1532, 'G', 1533-1534, 'G', 1535-1536, 'G', 1537-1538, 'G', 1539-1540, 'G', 1541-1542, 'G', 1543-1544, 'G', 1545-1546, 'G', 1547-1548, 'G', 1549-1550, 'G', 1551-1552, 'G', 1553-1554, 'G', 1555-1556, 'G', 1557-1558, 'G', 1559-1560, 'G', 1561-1562, 'G', 1563-1564, 'G', 1565-1566, 'G', 1567-1568, 'G', 1569-1570, 'G', 1571-1572, 'G', 1573-1574, 'G', 1575-1576, 'G', 1577-1578, 'G', 1579-1580, 'G', 1581-1582, 'G', 1583-1584, 'G', 1585-1586, 'G', 1587-1588, 'G', 1589-1590, 'G', 1591-1592, 'G', 1593-1594, 'G', 1595-1596, 'G', 1597-1598, 'G', 1599-1600, 'G', 1601-1602, 'G', 1603-1604, 'G', 1605-1606, 'G', 1607-1608, 'G', 1609-1610, 'G', 1611-1612, 'G', 1613-1614, 'G', 1615-1616, 'G', 1617-1618, 'G', 1619-1620, 'G', 1621-1622, 'G', 1623-1624, 'G', 1625-1626, 'G', 1627-1628, 'G', 1629-1630, 'G', 1631-1632, 'G', 1633-1634, 'G', 1635-1636, 'G', 1637-1638, 'G', 1639-1640, 'G', 1641-1642, 'G', 1643-1644, 'G', 1645-1646, 'G', 1647-1648, 'G', 1649-1650, 'G', 1651-1652, 'G', 1653-1654, 'G', 1655-1656, 'G', 1657-1658, 'G', 1659-1660, 'G', 1661-1662, 'G', 1663-1664, 'G', 1665-1666, 'G', 1667-1668, 'G', 1669-1670, 'G', 1671-1672, 'G', 1673-1674, 'G', 1675-1676, 'G', 1677-1678, 'G', 1679-1680, 'G', 1681-1682, 'G', 1683-1684, 'G', 1685-1686, 'G', 1687-1688, 'G', 1689-1690, 'G', 1691-1692, 'G', 1693-1694, 'G', 1695-1696, 'G', 1697-1698, 'G', 1699-1700, 'G', 1701-1702, 'G', 1703-1704, 'G', 1705-1706, 'G', 1707-1708, 'G', 1709-1710, 'G', 1711-1712, 'G', 1713-1714, 'G', 1715-1716, 'G', 1717-1718, 'G', 1719-1720, 'G', 1721-1722, 'G', 1723-1724, 'G', 1725-1726, 'G', 1727-1728, 'G', 1729-1730, 'G', 1731-1732, 'G', 1733-1734, 'G', 1735-1736, 'G', 1737-1738, 'G', 1739-1740, 'G', 1741-1742, 'G', 1743-1744, 'G', 1745-1746, 'G', 1747-1748, 'G', 1749-1750, 'G', 1751-1752, 'G', 1753-1754, 'G', 1755-1756, 'G', 1757-1758, 'G', 1759-1760, 'G', 1761-1762, 'G', 1763-1764, 'G', 1765-1766, 'G', 1767-1768, 'G', 1769-1770, 'G', 1771-1772, 'G', 1773-1774, 'G', 1775-1776, 'G', 1777-1778, 'G', 1779-1780, 'G', 1781-1782, 'G', 1783-1784, 'G', 1785-1786, 'G', 1787-1788, 'G', 1789-1790, 'G', 1791-1792, 'G', 1793-1794, 'G', 1795-1796, 'G', 1797-1798, 'G', 1799-1800, 'G', 1801-1802, 'G', 1803-1804, 'G', 1805-1806, 'G', 1807-1808, 'G', 1809-1810, 'G', 1811-1812, 'G', 1813-1814, 'G', 1815-1816, 'G', 1817-1818, 'G', 1819-1820, 'G', 1821-1822, 'G', 1823-1824, 'G', 1825-1826, 'G', 1827-1828, 'G', 1829-1830, 'G', 1831-1832, 'G', 1833-1834, 'G', 1835-1836, 'G', 1837-1838, 'G', 1839-1840, 'G', 1841-1842, 'G', 1843-1844, 'G', 1845-1846, 'G', 1847-1848, 'G', 1849-1850, 'G', 1851-1852, 'G', 1853-1854, 'G', 1855-1856, 'G', 1857-1858, 'G', 1859-1860, 'G', 1861-1862, 'G', 1863-1864, 'G', 1865-1866, 'G', 1867-1868, 'G', 1869-1870, 'G', 1871-1872, 'G', 1873-1874, 'G', 1875-1876, 'G', 1877-1878, 'G', 1879-1880, 'G', 1881-1882, 'G', 1883-1884, 'G', 1885-1886, 'G', 1887-1888, 'G', 1889-1890, 'G', 1891-1892, 'G', 1893-1894, 'G', 1895-1896, 'G', 1897-1898, 'G', 1899-1900, 'G', 1901-1902, 'G', 1903-1904, 'G', 1905-1906, 'G', 1907-1908, 'G', 1909-1910, 'G', 1911-1912, 'G', 1913-1914, 'G', 1915-1916, 'G', 1917-1918, 'G', 1919-1920, 'G', 1921-1922, 'G', 1923-1924, 'G', 1925-1926, 'G', 1927-1928, 'G', 1929-1930, 'G', 1931-1932, 'G', 1933-1934, 'G', 1935-1936, 'G', 1937-1938, 'G', 1939-1940, 'G', 1941-1942, 'G', 1943-1944, 'G', 1945-1946, 'G', 1947-1948, 'G', 1949-1950, 'G', 1951-1952, 'G', 1953-1954, 'G', 1955-1956, 'G', 1957-1958, 'G', 1959-1960, 'G', 1961-1962, 'G', 1963-1964, 'G', 1965-1966, 'G', 1967-1968, 'G', 1969-1970, 'G', 1971-1972, 'G', 1973-1974, 'G', 1975-1976, 'G', 1977-1978, 'G', 1979-1980, 'G', 1981-1982, 'G', 1983-1984, 'G', 1985-1986, 'G', 1987-1988, 'G', 1989-1990, 'G', 1991-1992, 'G', 1993-1994, 'G', 1995-1996, 'G', 1997-1998, 'G', 1999-2000, 'G', 2001-2002, 'G', 2003-2004, 'G', 2005-2006, 'G', 2007-2008, 'G', 2009-2010, 'G', 2011-2012, 'G', 2013-2014, 'G', 2015-2016, 'G', 2017-2018, 'G', 2019-2020, 'G', 2021-2022, 'G', 2023-2024, 'G', 2025-2026, 'G', 2027-2028, 'G', 2029-2030, 'G', 2031-2032, 'G', 2033-2034, 'G', 2035-2036, 'G', 2037-2038, 'G', 2039-2040, 'G', 2041-2042, 'G', 2043-2044, 'G', 2045-2046, 'G', 2047-2048, 'G', 2049-2050, 'G', 2051-2052, 'G', 2053-2054, 'G', 2055-2056, 'G', 2057-2058, 'G', 2059-2060, 'G', 2061-2062, 'G', 2063-2064, 'G', 2065-2066, 'G', 2067-2068, 'G', 2069-2070, 'G', 2071-2072, 'G', 2073-2074, 'G', 2075-2076, 'G', 2077-2078, 'G', 2079-2080, 'G', 2081-2082, 'G', 2083-2084, 'G', 2085-2086, 'G', 2087-2088, 'G', 2089-2090, 'G', 2091-2092, 'G', 2093-2094, 'G', 2095-2096, 'G

F:21-3381/Product: versican, splice form V0 #status predicted <MAT>  
F:57,331,352,817,965,1017,1333,1393,1437,1463,1653,1974,2045,2074,2103,2263,2290,2356,2d

Query Match 86.0%; Score 43; DB 2; Length 3381;  
Best Local Similarity 70.0%; Pred. No. 69;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
II |::|||:|

Db 3100 CVPYSGDRC 3109

RESULT 13

A40136  
fibropellin Ia - sea urchin (Strongylocentrotus purpuratus)  
N:Alternate names: epidermal growth factor homolog precursor  
N:Contains: alternatively spliced fibropellin Ib (EGF)  
C:Species: Strongylocentrotus purpuratus (purple urchin)  
C:Date: 13-May-1992 #sequence\_revision 17-Sep-1997 #text\_change 21-Jul-2000  
C:Accession: A40136; B40136; A29316; A43131  
R:Delgadillo-Reynoso, M.G.; Rollo, D.R.; Hursh, D.A.; Raff, R.A.  
J. Mol. Evol. 29, 314-327, 1989

A:Title: Structural analysis of the uEGF gene in the sea urchin Strongylocentrotus purpuratus  
A:Reference number: A40136; MUID:90112459; PMID:2514273

A:Accession: A40136  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-114 <DEL>  
A:Cross-references: GB:X17530; NID:910225; PID:9667061

A:Accession: B40136  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 181-251,329-370,'R',372-408,'RA',411-441 <DEL>  
A:Accession: C40136  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 'K',747-821,898-978 <DEL>  
R:Hursh, D.A.; Andrews, M.E.; Raff, R.A.  
Science 237, 1487-1490, 1987

A:Title: A sea urchin gene encodes a polypeptide homologous to epidermal growth factor.  
A:Reference number: A29316; MUID:87319677; PMID:3498216

A:Accession: A29316  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 'S',280-481,786-1064 <HUR>  
A:Cross-references: GB:M17421; NID:9161474; PIDN:AAA30050.1; PID:g5522260  
R:Hunt, L.T.; Barker, W.C.  
FASEB J. 3, 1760-1764, 1989

A:Title: Avidin-like domain in an epidermal growth factor homolog from a sea urchin.  
A:Reference number: A43131; MUID:89196806; PMID:2784773

A:Contents: annotation  
C:Comment: EGF homology repeats 10-17 are spliced out in the short form (fibropellin Ib)  
C:Superfamily: C1r/C1s repeat homology <C1R>  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1064/Product: fibropellin I #status predicted <FIB>  
F:23-54/Domain: EGF homology <EG01>  
F:57-175/Domain: C1r/C1s repeat homology <C1R>  
F:180-211/Domain: EGF homology <EG02>  
F:218-249/Domain: EGF homology <EG03>  
F:256-287/Domain: EGF homology <EG04>  
F:294-325/Domain: EGF homology <EG05>  
F:332-363/Domain: EGF homology <EG06>  
F:370-401/Domain: EGF homology <EG07>  
F:408-439/Domain: EGF homology <EG08>  
F:446-477/Domain: EGF homology <EG09>  
F:484-515/Domain: EGF homology <EG10>  
F:522-553/Domain: EGF homology <EG11>  
F:560-591/Domain: EGF homology <EG12>  
F:598-629/Domain: EGF homology <EG13>  
F:636-667/Domain: EGF homology <EG14>  
F:674-705/Domain: EGF homology <EG15>  
F:712-743/Domain: EGF homology <EG16>  
F:750-781/Domain: EGF homology <EG17>

F:788-819/Domain: EGF homology <EG18>  
F:826-857/Domain: EGF homology <EG19>  
F:864-895/Domain: EGF homology <EG20>  
F:902-933/Domain: EGF homology <EG21>  
F:936-1064/Region: avidin-like  
F:23-34,28-43,45-54,62-88,180-191,185-200,202-211,218-229,223-238,240-249,256-267,262  
57,451-466,468-477,484-495/Disulfide bonds: #status predicted  
F:489-504,506-515,522-533,527-542,544-553,560-571,565-580,582-591,598-609,603-618,62  
08,810-819,826-837,831-846,848-857,864-875,869-884,886-895,902-913,907-922,924-933/D

Query Match 84.0%; Score 42; DB 2; Length 1064;  
Best Local Similarity 60.0%; Pred. No. 34;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
I |::|||:|

Db 316 CPLGFGDNC 325

RESULT 14

T30201  
Notch homolog protein - sea squirt (Halocynthia roretzi)  
C:Species: Halocynthia roretzi  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T30201  
R:Hori, S.; Saitoh, T.; Matsumoto, M.; Makabe, K.W.; Nishida, H.  
Dev. Genes Evol. 207, 371-380, 1997

A:Title: Notch homologue from Halocynthia roretzi is preferentially expressed in the  
A:Reference number: Z20775  
A:Accession: T30201  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2352 <HOR>  
A:Cross-references: EMBL:AB001327; NID:d1204472; PID:d1026501; PIDN:BAA25571.1  
C:Genetics:  
A:Gene: Notch

Query Match 82.0%; Score 41; DB 2; Length 2352;  
Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
I |::|||:|

Db 443 CALGFTGDDC 452

RESULT 15

PC4415  
Erbb kinase activator beta, brain and thymus - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Dec-1997 #sequence\_revision 10-Dec-1997 #text\_change 02-Aug-2002  
C:Accession: PC4415  
R:Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.;  
J. Biochem. 122, 675-680, 1997  
A:Title: A novel brain-derived member of the epidermal growth factor family that int  
A:Reference number: JCS700; MUID:98006324; PMID:9348101  
A:Accession: PC4415  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-57 <HIG>  
A:Cross-references: DDBJ:D89997; NID:g2605633; PIDN:BAA23346.1; PID:g2605634  
A:Experimental source: PC-12 cell  
C:Comment: This protein is a member of the epidermal growth factor family. It is fun  
ating the differentiation of MDA-MB-453 cells.  
C:Superfamily: human Erbb kinase activator alpha, brain and thymus; EGF homology  
F:1-25/Domain: EGF homology (fragment) <EGF>

Query Match 80.0%; Score 40; DB 2; Length 57;  
Best Local Similarity 50.0%; Pred. No. 4.7;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10  
I |::|||:|

Wed Jul 2 08:08:18 2003

us-09-673-785d-13.jul1.rpr

Page 6

Db 16 CPVGTGDR 25

Search completed: July 2, 2003, 07:26:27  
Job time : 13 secs

2-100

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:02 ; Search time 10.3333 Seconds  
(without alignments)  
40.138 Million cell updates/sec

Title: US-09-673-785D-13  
Perfect score: 50  
Sequence: 1 CVIGXSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	50	100.0	1217	1 EGF_MOUSE	P01132 mus musculus
2	45	90.0	2139	1 CRB_DROME	P10040 drosophila
3	43	86.0	862	1 PGCV_MACNE	P28858 macaca nemo
4	43	86.0	2470	1 NTC2_MOUSE	O35516 mus musculus
5	43	86.0	2471	1 NTC2_HUMAN	O04721 homo sapien
6	43	86.0	2471	1 NTC2_RAT	O9QW30 rattus norv
7	43	86.0	3381	1 PGCV_BOVIN	P81282 bos taurus
8	43	86.0	3396	1 PGCV_HUMAN	P13611 homo sapien
9	42	84.0	1064	1 FBPI_STRPU	P10079 strongyloce
10	42	84.0	2920	1 CLR2_MOUSE	Q9R0M0 mus musculus
11	40	80.0	161	1 7B2_XENLA	P18844 xenopus lae
12	40	80.0	756	1 NRG2_MOUSE	P56974 mus musculus
13	40	80.0	1133	1 EGF_RAT	P07522 rattus norv
14	40	80.0	2437	1 NTC1_BRARE	P46530 brachydanio
15	39	78.0	833	1 DL_DROME	P10041 drosophila
16	39	78.0	883	1 PGCB_RAT	P55068 rattus norv
17	39	78.0	1049	1 ADP1_YEAST	P25371 saccharomyc
18	39	78.0	1207	1 EGF_HUMAN	P01133 homo sapien
19	39	78.0	1213	1 JAG3_BRARE	Q9QY54 brachydanio
20	39	78.0	2319	1 NTC3_RAT	Q9R172 rattus norv
21	39	78.0	2738	1 PGCV_RAT	Q9ERB4 rattus norv
22	39	78.0	3358	1 PGCV_MOUSE	P62059 mus musculus
23	38	76.0	72	1 MT12_MYTED	P80247 mytilus edu
24	38	76.0	636	1 DKS_RALSO	O8XX95 ralstonia s
25	38	76.0	1242	1 JAG1_BRARE	Q9QY57 brachydanio
26	38	76.0	2476	1 ZAN_PIG	P28983 sus scrofa
27	38	76.0	4289	1 TENX_HUMAN	P22105 homo sapien
28	38	76.0	5255	1 BACA_BACLI	O68006 b bacitraci
29	37	74.0	108	1 RODL_NEUCR	P04571 neurospora
30	37	74.0	738	1 TRFM_HUMAN	P08582 homo sapien
31	37	74.0	738	1 TRFM_MOUSE	Q9R0R1 mus musculus
32	37	74.0	1295	1 GLP1_CAEEL	P13508 caenorhabdi
33	37	74.0	2321	1 NTC3_HUMAN	Q9UM47 homo sapien

RESULT 1				
EGF_MOUSE	34	37	74.0	2556
ID	EGF_MOUSE	STANDARD;	PRT;	1217 AA.
AC	P01132;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal growth factor].			
GN	EGF.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=8323630; PubMed=6602382;			
RA	Scott J., Urdea M., Quiroga M., Sanchez-Pescador R., Fong N.M.,			
RA	Selby M., Rutter W.J., Bell G.I.;			
RT	"Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins.";			
RT	Science 221:236-240(1983).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=83219309; PubMed=6304537;			
RX	Gray A., Dull T.J., Ullrich A.;			
RA	"Nucleotide sequence of epidermal growth factor cDNA predicts a			
RT	128,000-molecular weight protein precursor.";			
RL	Nature 303:722-725(1983).			
RN	[3]			
RP	SEQUENCE OF 977-1029.			
RX	MEDLINE=73048516; PubMed=4636327;			
RA	Savage C.R. Jr., Inagami T., Cohen S.;			
RT	"The primary structure of epidermal growth factor.";			
RL	J. Biol. Chem. 247:7612-7621(1972).			
RN	[4]			
RP	DISULFIDE BONDS.			
RX	MEDLINE=74025498; PubMed=4750422;			
RA	Savage C.R. Jr., Hash J.H., Cohen S.;			
RT	"Epidermal growth factor. Location of disulfide bonds.";			
RL	J. Biol. Chem. 248:7669-7672(1973).			
RN	[5]			
RP	STRUCTURE BY NMR OF 977-1029.			
RX	MEDLINE=92118798; PubMed=1731873;			
RA	Montellone G.T., Wuehrich K., Burgess A.W., Nice E.C., Wagner G.,			
RT	Gibson K.D., Scheraga H.A.;			
RT	"Solution structure of murine epidermal growth factor determined by			
RT	NMR spectroscopy and refined by energy minimization with			
RT	restraints.";			
RN	[6]			
RP	Biochemistry 31:236-249(1992).			
RX	STRUCTURE BY NMR OF 977-1029.			
RX	MEDLINE=93075811; PubMed=1445923;			
RA	Kohda D., Inagaki F.;			
RT	"Three-dimensional nuclear magnetic resonance structures of mouse			
RT	epidermal growth factor in acidic and physiological pH solutions.";			

P46531 homo sapien  
Q21313 caenorhabdi  
Q00968 sus scrofa  
Q29845 archaeoglob  
O14944 homo sapien  
P29058 hevea bras  
Q25464 mytilus gal  
P55028 carassius a  
P43256 arabidopsis  
P49013 strongyloce  
P29057 hevea bras  
O64966 gossypium h

ALIGNMENTS

Biochemistry 31:11928-11939(1992).  
 [7]  
 RN STRUCTURE BY NMR OF 980-1024.  
 RX MEDLINE-99180407; PubMed-10082370;  
 RA Barnham K.J., Torres A.M., Alewood D., Alewood P.F., Domagala T.,  
 RA Nice E.C., Norton R.S.;  
 RT "Role of the 6-20 disulfide bridge in the structure and activity of  
 RT epidermal growth factor.";  
 RL Protein Sci. 7:1738-1749(1998).  
 CC -1- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS  
 CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME  
 CC FIBROBLASTS IN CELL CULTURE.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 1134  
 CC TO 1168 DUE TO A FRAMESHIFT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: J00380; AAA37539.1; -  
 CC EMBL: V00741; CAA24115.1; ALT\_FRAME.  
 CC EMBL: V00741; CAA24116.1; -  
 CC PIR: A01387; EGM5MG.  
 CC PDB: 1EGF; 31-JAN-94.  
 CC PDB: 3EGF; 31-JAN-94.  
 CC PDB: 1EPG; 31-JAN-94.  
 CC PDB: 1EPH; 31-JAN-94.  
 CC PDB: 1EPJ; 31-JAN-94.  
 CC PDB: 1EPJ; 31-JAN-94.  
 CC PDB: 1A3P; 29-JUL-98.  
 CC MGD: MGI:95290; Egf.  
 CC InterPro: IPR000152; Asx\_hydroxyl.  
 CC InterPro: IPR000561; EGF-like.  
 CC InterPro: IPR001336; EGF\_L.  
 CC InterPro: IPR001861; EGF\_Ca.  
 CC InterPro: IPR000033; Ldl\_receptor\_rep.  
 CC Pfam: PF00008; EGF; 8.  
 CC Pfam: PF00058; ldl\_recept\_b; 7.  
 CC PRINTS: PR00009; EGF\_TGF.  
 CC SMART: SM00179; EGF\_CA; 2.  
 CC SMART: SM00001; EGF\_like; 7.  
 CC SMART: SM00135; Ly; 9.  
 CC PROSITE: PS00010; ASX\_HYDROXYL; 3.  
 CC PROSITE: PS00022; EGF\_1; 1.  
 CC PROSITE: PS01186; EGF\_2; 6.  
 CC PROSITE: PS01187; EGF\_CA; 3.  
 CC EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;  
 KW Signal; 3D-structure.  
 FT SIGNAL 1 28  
 FT CHAIN 29 1217  
 FT PRO-EPIDERMAL GROWTH FACTOR.  
 FT CHAIN 977 1029  
 FT EPIDERMAL GROWTH FACTOR.  
 FT DOMAIN 29 1038  
 FT EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1039 1058  
 FT POTENTIAL.  
 FT DOMAIN 1059 1217  
 FT CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 327 361  
 FT EGF-LIKE 1 (INCOMPLETE).  
 FT DOMAIN 362 402  
 FT EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 403 443  
 FT EGF-LIKE 3.  
 FT DOMAIN 441 483  
 FT EGF-LIKE 4.  
 FT DOMAIN 747 787  
 FT EGF-LIKE 5.  
 FT DOMAIN 838 876  
 FT EGF-LIKE 6.  
 FT DOMAIN 877 918  
 FT EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 919 959  
 FT EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 978 1019  
 FT EGF-LIKE 9.  
 FT DISULFID 366 377  
 FT BY SIMILARITY.  
 FT DISULFID 373 386  
 FT BY SIMILARITY.  
 FT DISULFID 388 401  
 FT BY SIMILARITY.  
 FT DISULFID 407 418  
 FT BY SIMILARITY.

FT DISULFID 414 427 BY SIMILARITY.  
 FT DISULFID 429 442 BY SIMILARITY.  
 FT DISULFID 445 457 BY SIMILARITY.  
 FT DISULFID 453 467 BY SIMILARITY.  
 FT DISULFID 469 482 BY SIMILARITY.  
 FT DISULFID 751 762 BY SIMILARITY.  
 FT DISULFID 758 771 BY SIMILARITY.  
 FT DISULFID 773 786 BY SIMILARITY.  
 FT DISULFID 842 853 BY SIMILARITY.  
 FT DISULFID 847 862 BY SIMILARITY.  
 FT DISULFID 864 875 BY SIMILARITY.  
 FT DISULFID 881 895 BY SIMILARITY.  
 FT DISULFID 888 904 BY SIMILARITY.  
 FT DISULFID 906 917 BY SIMILARITY.  
 FT DISULFID 923 936 BY SIMILARITY.  
 FT DISULFID 930 945 BY SIMILARITY.  
 FT DISULFID 947 958 BY SIMILARITY.  
 FT DISULFID 982 996 BY SIMILARITY.  
 FT DISULFID 990 1007 BY SIMILARITY.  
 FT DISULFID 1009 1018 BY SIMILARITY.  
 FT DOMAIN 1024 1029 BY SIMILARITY.  
 FT CARBOHYD 111 111 NOT REQUIRED FOR FULL BIOLOGICAL ACTIVITY  
 FT CARBOHYD 410 410 IN VIVO.  
 FT CARBOHYD 810 810 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 790 790 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 1048 1048 D -> Y (IN REF. 2).  
 FT STRAND 995 997 A -> S (IN REF. 2).  
 FT STRAND 1006 1008  
 FT STRAND 1010 1010  
 FT TURN 1011 1012  
 FT TURN 1013 1014  
 FT STRAND 1020 1021  
 SQ SEQUENCE 1217 AA; 133144 MW; A9C7F3D512F82873 CRC64;  
 Query Match 100.0%; Score 50; DB 1; Length 1217;  
 Best Local Similarity 80.0%; Pred. No. 0.38;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGSGDXC 10  
 Db 1009 CVIGSGDRC 1018  
 III:III:II  
 RESULT 2  
 ID CRB\_DROME STANDARD; PRT; 2139 AA.  
 AC P10040;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Crumbs protein precursor (95F).  
 GN CRB.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-Oregon-R; TISSUE-Embryo;  
 RX MEDLINE-90263104; PubMed-2344615;  
 RA Tepass U., Theres C., Knust E.;  
 RT "Crumbs encodes an EGF-like protein expressed on apical membranes of  
 RT Drosophila epithelial cells and required for organization of  
 RT epithelia.";  
 RL Cell 61:787-799(1990).  
 RN [2]  
 RN SEQUENCE OF 1663-1955 FROM N.A.  
 RC TISSUE-Embryo;  
 RX MEDLINE-87218537; PubMed-3107986;  
 RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,

RA Vaessin H., Campos-Ortega J.A.;  
 RT "EGF homologous sequences encoded in the genome of Drosophila  
 RL melanogaster", and their relation to neurogenic genes.";   
 CC EMBO J. 6:761-766(1987).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,  
 CC POLARITY. IT MAY ACT AS A SIGNAL.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).  
 CC -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS  
 CC -!- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M33753; AAA28428.1; ALT\_SEQ.  
 CC EMBL: X05144; CAA28793.1; -  
 CC PIR: B26637; B26637.  
 CC PIR: A35672; A35672.  
 CC HSP: P00740; 1EDM.  
 CC FlyBase: FBgn000368; crb.  
 CC InterPro: IPR000152; Asx\_hydroxyl.  
 CC InterPro: IPR000561; EGF-like.  
 CC InterPro: IPR000742; EGF\_2.  
 CC InterPro: IPR001881; EGF\_Ca.  
 CC InterPro: IPR001438; EGF\_II.  
 CC InterPro: IPR001791; Laminin\_G.  
 CC Pfam: PF00008; EGF\_26.  
 CC Pfam: PF00054; laminin\_G; 3.  
 CC PRINTS: PR00010; EGFBLD.  
 CC SMART: SM00179; EGF\_CA; 16.  
 CC SMART: SM00001; EGF-like; 16.  
 CC SMART: SM00282; LamG; 3.  
 CC PROSITE: PS00010; ASX\_HYDROXYL; 15.  
 CC PROSITE: PS00022; EGF\_1; 26.  
 CC PROSITE: PS01186; EGF\_2; 17.  
 CC PROSITE: PS01187; EGF\_CA; 12.  
 CC PROSITE: PS00025; LAM\_G\_DOMAIN; 3.  
 CC Differentiation; Repeat; EGF-like domain; Transmembrane;  
 KW Glycoprotein; Signal; Phosphorylation.  
 FT SIGNAL 1 90  
 FT CHAIN 91 2139 CROMBS PROTEIN.  
 FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 2085 2111 POTENTIAL.  
 FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 267 303 EGF-LIKE 1.  
 FT DOMAIN 306 343 EGF-LIKE 2.  
 FT DOMAIN 348 386 EGF-LIKE 3.  
 FT DOMAIN 388 425 EGF-LIKE 4.  
 FT DOMAIN 427 463 EGF-LIKE 5.  
 FT DOMAIN 464 500 EGF-LIKE 6.  
 FT DOMAIN 501 532 EGF-LIKE 7.  
 FT DOMAIN 545 581 EGF-LIKE 8.  
 FT DOMAIN 582 611 EGF-LIKE 9.  
 FT DOMAIN 609 646 EGF-LIKE 10.  
 FT DOMAIN 648 685 EGF-LIKE 11.  
 FT DOMAIN 687 723 EGF-LIKE 12.  
 FT DOMAIN 725 761 EGF-LIKE 13.  
 FT DOMAIN 763 800 EGF-LIKE 14.  
 FT DOMAIN 802 838 EGF-LIKE 15.  
 FT DOMAIN 840 902 EGF-LIKE 16.  
 FT DOMAIN 904 940 EGF-LIKE 17.  
 FT DOMAIN 942 978 EGF-LIKE 18.  
 FT DOMAIN 980 1021 EGF-LIKE 19.  
 FT DOMAIN 1023 1205 LAMININ G-LIKE 1.  
 FT DOMAIN 1207 1243 EGF-LIKE 20.  
 FT DOMAIN 1250 1480 LAMININ G-LIKE 2.  
 FT DOMAIN 1481 1517 EGF-LIKE 21.

FT	DOMAIN	1558	1758	LAMININ G-LIKE 3.
FT	DOMAIN	1759	1795	EGF-LIKE 22.
FT	DOMAIN	1797	1833	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1835	1871	EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1874	1915	EGF-LIKE 25.
FT	DOMAIN	1915	1951	EGF-LIKE 26.
FT	DOMAIN	1953	1989	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1991	2029	EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	2030	2070	EGF-LIKE 29.
FT	DISULFID	271	282	BY SIMILARITY.
FT	DISULFID	276	291	BY SIMILARITY.
FT	DISULFID	293	302	BY SIMILARITY.
FT	DISULFID	310	321	BY SIMILARITY.
FT	DISULFID	315	331	BY SIMILARITY.
FT	DISULFID	333	342	BY SIMILARITY.
FT	DISULFID	352	363	BY SIMILARITY.
FT	DISULFID	357	374	BY SIMILARITY.
FT	DISULFID	376	385	BY SIMILARITY.
FT	DISULFID	392	403	BY SIMILARITY.
FT	DISULFID	397	412	BY SIMILARITY.
FT	DISULFID	414	424	BY SIMILARITY.
FT	DISULFID	431	442	BY SIMILARITY.
FT	DISULFID	436	451	BY SIMILARITY.
FT	DISULFID	453	462	BY SIMILARITY.
FT	DISULFID	468	479	BY SIMILARITY.
FT	DISULFID	473	488	BY SIMILARITY.
FT	DISULFID	490	499	BY SIMILARITY.
FT	DISULFID	505	515	BY SIMILARITY.
FT	DISULFID	509	520	BY SIMILARITY.
FT	DISULFID	522	531	BY SIMILARITY.
FT	DISULFID	549	562	BY SIMILARITY.
FT	DISULFID	556	569	BY SIMILARITY.
FT	DISULFID	571	580	BY SIMILARITY.
FT	DISULFID	586	597	BY SIMILARITY.
FT	DISULFID	591	602	BY SIMILARITY.
FT	DISULFID	604	610	BY SIMILARITY.
FT	DISULFID	613	624	BY SIMILARITY.
FT	DISULFID	618	634	BY SIMILARITY.
FT	DISULFID	636	645	BY SIMILARITY.
FT	DISULFID	652	664	BY SIMILARITY.
FT	DISULFID	659	673	BY SIMILARITY.
FT	DISULFID	675	684	BY SIMILARITY.
FT	DISULFID	691	702	BY SIMILARITY.
FT	DISULFID	696	711	BY SIMILARITY.
FT	DISULFID	713	722	BY SIMILARITY.
FT	DISULFID	729	740	BY SIMILARITY.
FT	DISULFID	734	749	BY SIMILARITY.
FT	DISULFID	751	760	BY SIMILARITY.
FT	DISULFID	767	778	BY SIMILARITY.
FT	DISULFID	772	787	BY SIMILARITY.
FT	DISULFID	789	799	BY SIMILARITY.
FT	DISULFID	806	817	BY SIMILARITY.
FT	DISULFID	811	826	BY SIMILARITY.
FT	DISULFID	828	837	BY SIMILARITY.
FT	DISULFID	844	855	BY SIMILARITY.
FT	DISULFID	849	890	BY SIMILARITY.
FT	DISULFID	892	901	BY SIMILARITY.
FT	DISULFID	908	919	BY SIMILARITY.
FT	DISULFID	913	928	BY SIMILARITY.
FT	DISULFID	930	939	BY SIMILARITY.
FT	DISULFID	946	957	BY SIMILARITY.
FT	DISULFID	952	966	BY SIMILARITY.
FT	DISULFID	968	977	BY SIMILARITY.
FT	DISULFID	984	995	BY SIMILARITY.
FT	DISULFID	989	1009	BY SIMILARITY.
FT	DISULFID	1011	1020	BY SIMILARITY.
FT	DISULFID	1211	1222	BY SIMILARITY.
FT	DISULFID	1216	1231	BY SIMILARITY.
FT	DISULFID	1233	1242	BY SIMILARITY.
FT	DISULFID	1485	1496	BY SIMILARITY.
FT	DISULFID	1490	1505	BY SIMILARITY.
FT	DISULFID	1507	1516	BY SIMILARITY.
FT	DISULFID	1763	1774	BY SIMILARITY.

FT DISULFID 1768 1783 BY SIMILARITY.  
 FT DISULFID 1785 1794 BY SIMILARITY.  
 FT DISULFID 1801 1812 BY SIMILARITY.  
 FT DISULFID 1806 1821 BY SIMILARITY.  
 FT DISULFID 1823 1832 BY SIMILARITY.  
 FT DISULFID 1839 1850 BY SIMILARITY.  
 FT DISULFID 1844 1859 BY SIMILARITY.  
 FT DISULFID 1861 1870 BY SIMILARITY.  
 FT DISULFID 1878 1889 BY SIMILARITY.  
 FT DISULFID 1883 1903 BY SIMILARITY.  
 FT DISULFID 1905 1914 BY SIMILARITY.  
 FT DISULFID 1919 1930 BY SIMILARITY.  
 FT DISULFID 1924 1939 BY SIMILARITY.  
 FT DISULFID 1941 1950 BY SIMILARITY.  
 FT DISULFID 1957 1968 BY SIMILARITY.  
 FT DISULFID 1962 1977 BY SIMILARITY.  
 FT DISULFID 1979 1988 BY SIMILARITY.  
 FT DISULFID 1995 2008 BY SIMILARITY.  
 FT DISULFID 2002 2017 BY SIMILARITY.  
 FT DISULFID 2019 2028 BY SIMILARITY.  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 Query Match Score 45; DB 1; Length 2139;  
 Best Local Similarity 60.0%; Pred. No. 5.2;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDC 10  
 Db 571 CAVGYSGDC 580  
 RESULT 3  
 PGCV\_MACNE  
 ID PGCV\_MACNE STANDARD; PRT; 862 AA.  
 AC Q28858; Q28859; Q28860;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Versican core protein (large fibroblast proteoglycan) (Chondroitin  
 sulfate proteoglycan core protein 2) (Fragments).  
 GN CSF62.  
 OS Macaca nemestrina (Pig-tailed macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 OX NCBI\_TaxID=9545;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Aortic smooth muscle;  
 RX MEDLINE=95005762; PubMed=7921538;  
 RA Yao L.Y., Moody C., Schoenherr E., Wight T.N., Sandell L.J.;  
 RT "Identification of the proteoglycan versican in aorta and smooth  
 muscle cells by DNA sequence analysis, in situ hybridization and  
 immunohistochemistry.";  
 RL Matrix Biol. 14:213-225(1994).  
 CC -!- FUNCTION: May play a role in intercellular signaling and in  
 connecting cells with the extracellular matrix. May take part in  
 the regulation of cell motility, growth and differentiation. Binds  
 hyaluronan.  
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By  
 similarity).  
 CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development  
 (By similarity).  
 CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.

CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; S72412; AAA65593.2; -;  
 CC EMBL; S72413; AAA65594.2; -;  
 CC EMBL; S72414; AAA65595.2; -;  
 CC HSP; P01132; EGF.  
 DR InterPro; IPR00152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_CA.  
 DR InterPro; IPR001438; EGF\_II.  
 DR InterPro; IPR001304; Lectin\_C.  
 DR InterPro; IPR000538; Link.  
 DR PRINTS; PR00010; EGFBLD.  
 DR PRODOM; PD000918; Link; 2.  
 DR SMART; SM00034; CLECT; 1.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00001; EGF-like; 1.  
 DR SMART; SM00445; LINK; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS01241; LINK; 1.  
 DR PROSITE; PS00615; C-TYPE LECTIN\_1; PARTIAL.  
 DR PROSITE; PS00041; C-TYPE LECTIN\_2; 1.  
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Repeat;  
 KW EGF-like domain; Calcium.  
 FT NON\_TER 1  
 FT DOMAIN <1 37 LINK 1.  
 FT DOMAIN 58 139 LINK 2.  
 FT DOMAIN 141 >233 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT  
 DOMAIN).  
 FT NON\_CONS 233 234 GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN,  
 DOMAIN <234 >525 SIMILAR TO CHONDROITIN SULFATE ATTACHMENT  
 SITE IN COLLAGEN TYPE IX (BY SIMILARITY).  
 FT NON\_CONS 525 526 EGF-LIKE 1.  
 FT DOMAIN 718 754 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 756 792 C-TYPE LECTIN.  
 FT DOMAIN 805 >862 BY SIMILARITY.  
 FT DISULFID 63 138 BY SIMILARITY.  
 FT DISULFID 87 108 BY SIMILARITY.  
 FT DISULFID 722 733 BY SIMILARITY.  
 FT DISULFID 727 742 BY SIMILARITY.  
 FT DISULFID 744 753 BY SIMILARITY.  
 FT DISULFID 760 771 BY SIMILARITY.  
 FT DISULFID 765 780 BY SIMILARITY.  
 FT DISULFID 782 791 BY SIMILARITY.  
 FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 563 563 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 696 696 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT NON\_TER 862 862  
 SQ SEQUENCE 862 AA; 95583 MW; A5D5F6153A74BB39 CRC64;  
 Query Match Score 43; DB 1; Length 862;  
 Best Local Similarity 70.0%; Pred. No. 4.9;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDC 10



DB 744 CVPGYSGDC 753

II | : | | : |

RESULT 4

NTC2\_MOUSE STANDARD; PRT: 2470 AA.

AC O35516; Q60941; Q06008;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Motch  
B).

DE NOTCH2.

GN Mus musculus (Mouse).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Thymus;

RA Hamada Y., Higuchi M., Tsujimoto Y.;

RT "Complete amino acid sequence and multiform transcripts encoded by a  
RT single copy of mouse Notch2 gene.";

RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 316-1518 FROM N.A.

RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;

RX MEDLINE=93178563; PubMed=8440332;

RA Lardelli M., Lendahl U.;

RT "Motch A and Motch B-two mouse Notch homologues coexpressed in a  
RT wide variety of tissues.";

RL Exp. Cell Res. 204:364-372(1993).

RN [3]

RP SEQUENCE OF 1765-2153 FROM N.A.

RX MEDLINE=97075110; PubMed=8917536;

RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,  
RA Martin D.I.;

RT "Inhibition of granulocytic differentiation by mNotch1.";

RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).

RN [4]

RP FUNCTION.

RX MEDLINE=99396706; PubMed=10393120;

RA Hamada Y., Kadohawa Y., Okabe M., Ikawa M., Coleman J.R.,  
RA Tsujimoto Y.;

RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early  
RT embryonic lethality.";

RL Development 126:3415-3424(1999).

RN [5]

RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.

RX MEDLINE=95333893; PubMed=7609614;

RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;

RT "Differential expression of Notch1 and Notch2 in developing and adult  
RT mouse brain.";

RL Brain Res. Mol. Brain Res. 29:263-272(1995).

RN [6]

RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.

RX MEDLINE=21523956; PubMed=11518718;

RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;

RT "Murine notch homologs (N1-4) undergo presenilin-dependent  
RT proteolysis.";

RL J. Biol. Chem. 276:40268-40273(2001).

RN [7]

RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.

RX MEDLINE=21374376; PubMed=11459941;

RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;

RT "Conservation of the biochemical mechanisms of signal transduction  
RT among mammalian Notch family members.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).

CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands  
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.  
CC Upon ligand activation through the released notch intracellular  
CC domain (NICD) it forms a transcriptional activator complex with

CC RBP-J kappa and activates genes of the enhancer of split locus.  
CC Affects the implementation of differentiation, proliferation and  
CC apoptotic programs (By similarity). May play an essential role in  
CC postimplantation development, probably in some aspect of cell  
CC specification and/or differentiation.  
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
CC terminal fragment N(EC) which are probably linked by disulfide  
CC bonds.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following  
CC proteolytical processing NICD is translocated to the nucleus.  
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
CC produced by alternative splicing.  
CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,  
CC neuroepithelia, somites, optic vesicles and branchial arches, but  
CC not heart.  
CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,  
CC the postnatal ependymal cells, and the choroid plexus throughout  
CC embryonic and postnatal development.  
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
CC which is proteolytically cleaved by a furin-like convertase in the  
CC trans-Golgi network before it reaches the plasma membrane to yield  
CC an active, ligand-accessible form. Cleavage results in a C-  
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
CC ligand binding, it is cleaved by TNF-alpha converting enzyme  
CC (TACE) to yield a membrane-associated intermediate fragment called  
CC notch extracellular truncation (NEXT). This fragment is then  
CC cleaved by presenilin dependent gamma-secretase to release a  
CC notch-derived peptide containing the intracellular domain (NICD)  
CC from the membrane.  
CC -1- PTM: Phosphorylated.  
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.  
CC -1- SIMILARITY: CONTAINS 34.5 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.  
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: D32210; BAA22094.1; -  
CC EMBL: G68279; CAA48340.1; -  
CC EMBL: U31881; AAC52924.1; -  
CC HSPSP: P16109; IFSB.  
CC MGD: MGI:97364; Notch2.  
CC InterPro: IPR002110; ANK.  
CC InterPro: IPR000152; Asx\_hydroxyl.  
CC InterPro: IPR000561; EGF-like.  
CC InterPro: IPR000742; EGF\_2.  
CC InterPro: IPR001881; EGF\_Ca.  
CC InterPro: IPR001438; EGF\_II.  
CC InterPro: IPR000800; Notch.  
CC Pfam: PF00008; EGF; 35.  
CC Pfam: PF00023; ank; 6.  
CC Pfam: PF00066; notch; 2.  
CC PRINTS: PR00010; EGFBL00D.  
CC PRINTS: PR01452; NOTCH.  
CC SMART: SM00248; ANK; 4.  
CC SMART: SM00179; EGF\_CA; 22.  
CC SMART: SM00001; EGF\_like; 12.  
CC SMART: SM00004; NL; 3.  
CC PROSITE: PS50088; ANK\_REPEAT; 4.  
CC PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
CC PROSITE: PS00010; ASX\_HYDROXYL; 22.  
CC PROSITE: PS00022; EGF\_1; 33.  
CC PROSITE: PS01186; EGF\_2; 27.  
CC PROSITE: PS01187; EGF\_CA; 22.  
KW Receptor; Transcription regulation; Activator; Differentiation;  
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;  
KW Transmembrane; Glycoprotein; Signal; Phosphorylation;



notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (by similarity).

-!- PTM: Phosphorylated (By similarity).

-!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.

-!- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.

-!- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.

-!- SIMILARITY: CONTAINS 6 ANK REPEATS.

-----

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AF308601; AAA36377.2; -.

EMBL; AF315356; AAG37073.1; -.

EMBL; U77493; AAB19224.1; -.

HSSP; P00740; 1EDM.

Genew; HGNC:7882; NOTCH2.

MIM; 600275; -.

InterPro; IPR002110; ANK.

InterPro; IPR000152; Asx\_hydroxyl.

InterPro; IPR000561; EGF-like.

InterPro; IPR000742; EGF-2.

InterPro; IPR001881; EGF-Ca.

InterPro; IPR001438; EGF-II.

InterPro; IPR002049; Laminin\_EGF.

InterPro; IPR000800; Notch.

Pfam; PF00008; EGF; 35.

Pfam; PF00023; ank; 6.

Pfam; PF00066; notch; 2.

PRINTS; PR00010; EGF\_BLOOD.

PRINTS; PR00011; EGF\_LAMININ.

SMART; SM00248; ANK; 4.

SMART; SM00179; EGF\_Ca; 22.

SMART; SM00001; EGF-like; 12.

SMART; SM00004; ND; 2.

PROSITE; PS50088; ANK\_REPEAT; 4.

PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.

PROSITE; PS00010; ASX\_HYDROXYL; 22.

PROSITE; PS00022; EGF\_1; 34.

PROSITE; PS01186; EGF\_2; 29.

PROSITE; PS01187; EGF\_Ca; 22.

Receptor; Transcription regulation; Activator; Differentiation; Developmental protein; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation.

SIGNAL 1 25

CHAIN 26 2471

CHAIN 1666 2471

CHAIN 1697 2471

DOMAIN 26 1677

TRANSMEM 1678 1698

DOMAIN 1699 2471

DOMAIN 26 63

DOMAIN 64 102

DOMAIN 105 143

DOMAIN 144 180

DOMAIN 182 219

DOMAIN 221 238

DOMAIN 260 296

DOMAIN 298 336

DOMAIN 338 374

DOMAIN 375 413

DOMAIN 415 454

DOMAIN 456 492

DOMAIN 494 530

DOMAIN 532 568

notch intracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (by similarity).

-!- PTM: Phosphorylated (By similarity).

-!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.

-!- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.

-!- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.

-!- SIMILARITY: CONTAINS 6 ANK REPEATS.

-----

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AF308601; AAA36377.2; -.

EMBL; AF315356; AAG37073.1; -.

EMBL; U77493; AAB19224.1; -.

HSSP; P00740; 1EDM.

Genew; HGNC:7882; NOTCH2.

MIM; 600275; -.

InterPro; IPR002110; ANK.

InterPro; IPR000152; Asx\_hydroxyl.

InterPro; IPR000561; EGF-like.

InterPro; IPR000742; EGF-2.

InterPro; IPR001881; EGF-Ca.

InterPro; IPR001438; EGF-II.

InterPro; IPR002049; Laminin\_EGF.

InterPro; IPR000800; Notch.

Pfam; PF00008; EGF; 35.

Pfam; PF00023; ank; 6.

Pfam; PF00066; notch; 2.

PRINTS; PR00010; EGF\_BLOOD.

PRINTS; PR00011; EGF\_LAMININ.

SMART; SM00248; ANK; 4.

SMART; SM00179; EGF\_Ca; 22.

SMART; SM00001; EGF-like; 12.

SMART; SM00004; ND; 2.

PROSITE; PS50088; ANK\_REPEAT; 4.

PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.

PROSITE; PS00010; ASX\_HYDROXYL; 22.

PROSITE; PS00022; EGF\_1; 34.

PROSITE; PS01186; EGF\_2; 29.

PROSITE; PS01187; EGF\_Ca; 22.

Receptor; Transcription regulation; Activator; Differentiation; Developmental protein; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation.

SIGNAL 1 25

CHAIN 26 2471

CHAIN 1666 2471

CHAIN 1697 2471

DOMAIN 26 1677

TRANSMEM 1678 1698

DOMAIN 1699 2471

DOMAIN 26 63

DOMAIN 64 102

DOMAIN 105 143

DOMAIN 144 180

DOMAIN 182 219

DOMAIN 221 238

DOMAIN 260 296

DOMAIN 298 336

DOMAIN 338 374

DOMAIN 375 413

DOMAIN 415 454

DOMAIN 456 492

DOMAIN 494 530

DOMAIN 532 568

FT	DOMAIN	570	605	EGF-LIKE 15,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	607	643	EGF-LIKE 16,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	645	680	EGF-LIKE 17,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	682	718	EGF-LIKE 18,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	720	755	EGF-LIKE 19.	
FT	DOMAIN	757	793	EGF-LIKE 20,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	795	831	EGF-LIKE 21,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	833	871	EGF-LIKE 22.	
FT	DOMAIN	873	909	EGF-LIKE 23,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	911	947	EGF-LIKE 24,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	949	985	EGF-LIKE 25,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	987	1023	EGF-LIKE 26,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1025	1061	EGF-LIKE 27,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1063	1099	EGF-LIKE 28.	
FT	DOMAIN	1101	1147	EGF-LIKE 29.	
FT	DOMAIN	1149	1185	EGF-LIKE 30,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1187	1223	EGF-LIKE 31,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1225	1262	EGF-LIKE 32,	
FT	DOMAIN	1264	1302	EGF-LIKE 33.	
FT	DOMAIN	1304	1343	EGF-LIKE 34.	
FT	DOMAIN	1374	1412	EGF-LIKE 35.	
FT	REPEAT	1420	1456	LIN/NOTCH 1.	
FT	REPEAT	1503	1535	LIN/NOTCH 2.	
FT	REPEAT	1827	1871	ANK 1.	
FT	REPEAT	1876	1905	ANK 2.	
FT	REPEAT	1909	1939	ANK 3.	
FT	REPEAT	1943	1972	ANK 4.	
FT	REPEAT	1976	2005	ANK 5.	
FT	REPEAT	2009	2038	ANK 6.	
FT	DOMAIN	1645	1648	POLY-ALA.	
FT	DOMAIN	1994	1997	POLY-LEU.	
FT	DOMAIN	2426	2429	POLY-SER.	
FT	DISULFID	28	41	BY SIMILARITY.	
FT	DISULFID	35	51	BY SIMILARITY.	
FT	DISULFID	53	62	BY SIMILARITY.	
FT	DISULFID	68	79	BY SIMILARITY.	
FT	DISULFID	73	90	BY SIMILARITY.	
FT	DISULFID	92	101	BY SIMILARITY.	
FT	DISULFID	109	121	BY SIMILARITY.	
FT	DISULFID	115	131	BY SIMILARITY.	
FT	DISULFID	133	142	BY SIMILARITY.	
FT	DISULFID	148	159	BY SIMILARITY.	
FT	DISULFID	153	168	BY SIMILARITY.	
FT	DISULFID	170	179	BY SIMILARITY.	
FT	DISULFID	186	198	BY SIMILARITY.	
FT	DISULFID	192	207	BY SIMILARITY.	
FT	DISULFID	209	218	BY SIMILARITY.	
FT	DISULFID	225	236	BY SIMILARITY.	
FT	DISULFID	230	246	BY SIMILARITY.	
FT	DISULFID	248	257	BY SIMILARITY.	
FT	DISULFID	264	275	BY SIMILARITY.	
FT	DISULFID	269	284	BY SIMILARITY.	
FT	DISULFID	286	295	BY SIMILARITY.	

Query Match 86.0%; Score 43; DB 1; Length 2471;

Best Local Similarity 70.0%; Pred. No. 14;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10

Db 326 CVNGWSGDDC 335

RESULT 6

NTC2\_RAT STANDARD; PRT; 2471 AA.

AC Q9QW30;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE 13-JUN-2002 (Rel. 41, Last annotation update)

DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).

GN NOTCH2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 FN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93202015; PubMed=1295745;  
 RA Weinmaster G., Roberts V.J., Lemke G.;  
 RL "Notch2: a second mammalian Notch gene.";  
 RT Development 116:931-941(1992).  
 RN [2]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=21331789; PubMed=11438922;  
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;  
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple  
 functional roles for the Notch-DSL signaling system during brain  
 development.";  
 RL J. Comp. Neurol. 436:167-181(2001).  
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands  
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.  
 CC Upon ligand activation through the released notch intracellular  
 CC domain (NICD) it forms a transcriptional activator complex with  
 CC RBP-J kappa and activates genes of the enhancer of split locus.  
 CC Affects the implementation of differentiation, proliferation and  
 CC apoptotic programs. May play an essential role in postimplantation  
 CC development, probably in some aspect of cell specification and/or  
 CC differentiation (By similarity).  
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
 CC terminal fragment N(IEC) which are probably linked by disulfide  
 CC bonds (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following  
 CC proteolytical processing NICD is translocated to the nucleus.  
 CC -!- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid  
 CC plexus in the brain. Expressed in postnatal central nervous system  
 CC (CNS) germinal zones and, in early postnatal life, within numerous  
 CC cells throughout the CNS. It is more highly localized to  
 CC ventricular germinal zones. Also found in the heart, liver and  
 CC kidney.  
 CC -!- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.  
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
 CC which is proteolytically cleaved by a furin-like convertase in the  
 CC trans-Golgi network before it reaches the plasma membrane to yield  
 CC an active, ligand-accessible form. Cleavage results in a C-  
 CC terminal fragment N(TM) and a N-terminal fragment N(IEC). Following  
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme  
 CC (TACE) to yield a membrane-associated intermediate fragment called  
 CC notch extracellular truncation (NEXT). This fragment is then  
 CC cleaved by presenilin dependent gamma-secretase to release a  
 CC notch-derived peptide containing the intracellular domain (NICD)  
 CC from the membrane (By similarity).  
 CC -!- PTM: Phosphorylated (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.  
 CC -!- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.  
 CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M93661; AAK13558.1; --  
 DR HSSP; P00743; ICCF.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF-2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001438; EGF\_II.  
 DR InterPro: IPR002049; Laminin\_EGF.

DR InterPro: IPR000800; Notch.  
 DR Pfam; PF00008; EGF; 35.  
 DR PFam; PF00023; ank; 6.  
 DR PFam; PF00066; notch; 2.  
 DR PRINTS; PR00010; EGFBLLOOD.  
 DR PRINTS; PR00011; EGF-LAMININ.  
 DR PRINTS; PR01452; NOTCH.  
 DR SMART; SM00248; ANK; 4.  
 DR SMART; SM00179; EGF\_CA; 22.  
 DR SMART; SM00001; EGF\_Like; 10.  
 DR SMART; SM00004; NL; 2.  
 DR PROSITE; PS00088; ANK\_REPEAT; 4.  
 DR PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE; PS00022; EGF\_1; 34.  
 DR PROSITE; PS01186; EGF\_2; 26.  
 DR PROSITE; PS01187; EGF\_CA; 22.  
 KW Receptor; Transcription regulation; Activator; Differentiation;  
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;  
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.  
 FT SIGNAL 1 25  
 FT CHAIN 26 2471  
 FT CHAIN 1666 2471  
 FT CHAIN 1697 2471  
 FT DOMAIN 26 1677  
 FT TRANSMEM 1678 1698  
 FT DOMAIN 1699 2471  
 FT DOMAIN 26 63  
 FT DOMAIN 64 102  
 FT DOMAIN 105 143  
 FT DOMAIN 144 180  
 FT DOMAIN 182 219  
 FT DOMAIN 221 258  
 FT DOMAIN 260 296  
 FT DOMAIN 298 336  
 FT DOMAIN 338 374  
 FT DOMAIN 375 413  
 FT DOMAIN 415 454  
 FT DOMAIN 456 492  
 FT DOMAIN 494 530  
 FT DOMAIN 532 568  
 FT DOMAIN 570 605  
 FT DOMAIN 607 643  
 FT DOMAIN 645 680  
 FT DOMAIN 682 718  
 FT DOMAIN 720 755  
 FT DOMAIN 757 793  
 FT DOMAIN 795 831  
 FT DOMAIN 833 871  
 FT DOMAIN 873 909  
 FT DOMAIN 911 947  
 FT DOMAIN 949 985  
 FT DOMAIN 987 1023  
 FT DOMAIN 1025 1061  
 FT DOMAIN 1063 1099  
 FT DOMAIN 1101 1147  
 FT DOMAIN 1149 1185  
 FT DOMAIN 1187 1223  
 FT DOMAIN 1225 1262  
 FT DOMAIN 1264 1302  
 FT DOMAIN 1304 1343  
 FT DOMAIN 1374 1412  
 FT DOMAIN 1645 1648  
 FT DOMAIN 1994 1997  
 FT DOMAIN 2426 2429  
 FT DOMAIN 2446 2451  
 FT REPEAT 1420 1456  
 FT REPEAT 1503 1535  
 FT REPEAT 1827 1871  
 FT REPEAT 1876 1905  
 FT REPEAT 1909 1939

```

FT REPEAT 1943 1972 ANK 4.
FT REPEAT 1976 2005 ANK 5.
FT REPEAT 2009 2038 ANK 6.
FT DISULFID 28 41 BY SIMILARITY.
FT DISULFID 35 51 BY SIMILARITY.
FT DISULFID 53 62 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 90 BY SIMILARITY.
FT DISULFID 92 101 BY SIMILARITY.
FT DISULFID 109 121 BY SIMILARITY.
FT DISULFID 115 131 BY SIMILARITY.
FT DISULFID 133 142 BY SIMILARITY.
FT DISULFID 148 159 BY SIMILARITY.
FT DISULFID 153 168 BY SIMILARITY.
FT DISULFID 170 179 BY SIMILARITY.
FT DISULFID 186 198 BY SIMILARITY.
FT DISULFID 192 207 BY SIMILARITY.
FT DISULFID 209 218 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 230 246 BY SIMILARITY.
FT DISULFID 248 257 BY SIMILARITY.
FT DISULFID 264 275 BY SIMILARITY.
FT DISULFID 289 284 BY SIMILARITY.
FT DISULFID 286 295 BY SIMILARITY.
FT DISULFID 302 315 BY SIMILARITY.
FT DISULFID 309 324 BY SIMILARITY.
FT DISULFID 326 335 BY SIMILARITY.
FT DISULFID 342 353 BY SIMILARITY.
FT DISULFID 347 362 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 379 390 BY SIMILARITY.
FT DISULFID 384 401 BY SIMILARITY.
FT DISULFID 403 412 BY SIMILARITY.
FT DISULFID 419 433 BY SIMILARITY.
FT DISULFID 427 442 BY SIMILARITY.
FT DISULFID 444 453 BY SIMILARITY.
FT DISULFID 460 471 BY SIMILARITY.
FT DISULFID 465 480 BY SIMILARITY.
FT DISULFID 482 491 BY SIMILARITY.
FT DISULFID 498 509 BY SIMILARITY.
FT DISULFID 503 518 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 536 547 BY SIMILARITY.
FT DISULFID 541 556 BY SIMILARITY.
FT DISULFID 558 567 BY SIMILARITY.
FT DISULFID 574 584 BY SIMILARITY.

Query Match 86.0%; Score 43; DB 1; Length 2471;
Best Local Similarity 70.0%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXGDXC 10
Db 326 CVNWSGDDC 335

RESULT 7
PCGV_BOVIN STANDARD; PRT: 3381 AA.
ID P81282; 077609; 077610; 077611; 077612;
AC 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glia1
DE hyaluronate-binding protein) (GHP).
GN CP5G2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]

```

```

RP SEQUENCE FROM N.A. (ISOFORMS V0; V1; V2 AND V3).
RC TISSUE=Forebrain;
RX MEDLINE=98288320; PubMed=9624174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain.";
RL J. Biol. Chem. 273:15758-15764(1998).
RN [2]
RP SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331
RP AND 342-348.
RC TISSUE=Spinal cord;
RX MEDLINE=92062692; PubMed=1720020;
RA Perides G., Biviano F., Bignami A.;
RT "Interaction of a brain extracellular matrix protein with hyaluronic
RT acid.";
RL Biochim. Biophys. Acta 1075:248-258(1991).
CC -1- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1, V2
CC and V3; are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed
CC in the central nervous system, and in a number of mesenchymal and
CC epithelial tissues; the major isoform V2 is restricted to the
CC central nervous system.
CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF060456; AAC24358.1; -
DR EMBL; AF060457; AAC24359.1; -
DR EMBL; AF060458; AAC24360.1; -
DR EMBL; AF060459; AAC24361.1; -
DR HSSP; P01132; IEPG.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003600; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR PRODOM; PD000918; Link; 1.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00409; IG; 1.
DR SMART; SM00410; IG_like; 1.

```

DR SMART; SMO0445; LINK; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.  
 DR PROSITE; PS00411; C-TYPE LECTIN 2; 1.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS01241; LINK; 2.  
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;  
 KW signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
 KW Hyaluronic acid; Alternative splicing.  
 FT SIGNAL 1 20  
 FT CHAIN 21 3381  
 FT DOMAIN 37 138  
 FT DOMAIN 168 245  
 FT DOMAIN 266 347  
 FT DOMAIN 349 1336  
 FT DOMAIN 1337 3074  
 FT DOMAIN 3074 3110  
 FT DOMAIN 3112 3148  
 FT DOMAIN 3161 3275  
 FT DOMAIN 3280 3338  
 FT DISULFID 44 131  
 FT DISULFID 173 244  
 FT DISULFID 197 218  
 FT DISULFID 271 346  
 FT DISULFID 295 316  
 FT DISULFID 3078 3089  
 FT DISULFID 3083 3098  
 FT DISULFID 3100 3109  
 FT DISULFID 3116 3127  
 FT DISULFID 3121 3136  
 FT DISULFID 3138 3147  
 FT DISULFID 3154 3165  
 FT DISULFID 3182 3274  
 FT DISULFID 3250 3266  
 FT DISULFID 3281 3324  
 FT DISULFID 3310 3337  
 FT CARBOHYD 57 57  
 FT CARBOHYD 331 331  
 FT CARBOHYD 352 352  
 FT CARBOHYD 817 817  
 FT CARBOHYD 965 965  
 FT CARBOHYD 1017 1017  
 FT CARBOHYD 1333 1333  
 FT CARBOHYD 1393 1393  
 FT CARBOHYD 1437 1437  
 FT CARBOHYD 1463 1463  
 FT CARBOHYD 1653 1653  
 FT CARBOHYD 1974 1974  
 FT CARBOHYD 2045 2045  
 FT CARBOHYD 2074 2074  
 FT CARBOHYD 2103 2103  
 FT CARBOHYD 2263 2263  
 FT CARBOHYD 2290 2290  
 FT CARBOHYD 2356 2356  
 FT CARBOHYD 2623 2623  
 FT CARBOHYD 2641 2641  
 FT CARBOHYD 2919 2919  
 FT CARBOHYD 3052 3052  
 FT CARBOHYD 3354 3354  
 FT CARBOHYD 3364 3364  
 FT VARSPLIC 349 349  
 FT VARSPLIC 350 1336  
 FT VARSPLIC 1337 3074  
 FT VARSPLIC 350 3074  
 FT CONFLICT 25 25  
 FT CONFLICT 51 51  
 FT CONFLICT 89 89  
 FT CONFLICT 96 96  
 FT CONFLICT 346 346  
 FT CONFLICT 3381 AA; 369984 MW; F09716FA7778D459 CRC64;  
 SQ SEQUENCE

Query Match 86.0%; Score 43; DB 1; Length 3381;  
 Best Local Similarity 70.0%; Pred. NO. 19;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CVIGSGDXC 10  
 Db 3100 CVPFGSDRC 3109  
 |||:|:|:  
 |||:|:|:  
 RESULT 8  
 PGCV\_HUMAN  
 ID PGCV\_HUMAN STANDARD; PRT: 3396 AA.  
 AC P13611; P20754; Q9UNW5; Q13010; Q13189; Q15123;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Versican core protein precursor (Large fibroblast proteoglycan)  
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glia)  
 DE hyaluronate-binding protein) (GHP).  
 GN CSPG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM V0).  
 RX MEDLINE=95105188; PubMed=7528742;  
 RA Naso M.F., Zimmermann D.R., Iozzo R.V.;  
 RT "Characterization of the complete genomic structure of the human  
 RT versican gene and functional analysis of its promoter.";  
 RL J. Biol. Chem. 269:32999-33008(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM V1).  
 RC TISSUE=Placenta;  
 RX MEDLINE=90059882; PubMed=2583089;  
 RA Zimmermann D.R., Ruoslahti E.;  
 RT "Multiple domains of the large fibroblast proteoglycan, versican.";  
 RL EMBO J. 8:2975-2981(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM V2).  
 RC TISSUE=Glial tumor;  
 RX MEDLINE=95105187; PubMed=7806529;  
 RA Dours-Zimmermann M.T., Zimmermann D.R.;  
 RT "A novel glycosaminoglycan attachment domain identified in two  
 RT alternative splice variants of human versican.";  
 RL J. Biol. Chem. 269:32992-32998(1994).  
 RN [4]  
 RP SEQUENCE OF 2711-3396 FROM N.A.  
 RC TISSUE=Lung fibroblast;  
 RX MEDLINE=88007514; PubMed=2820964;  
 RA Krusius T., Gehlsen K.R., Ruoslahti E.;  
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains  
 RT lectin-like and growth factor-like sequences.";  
 RL J. Biol. Chem. 262:13120-13125(1987).  
 RN [5]  
 RP SEQUENCE OF 251-347 FROM N.A.  
 RX MEDLINE=93122792; PubMed=1478664;  
 RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,  
 RA McPherson J.D.;  
 RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of  
 RT human chromosome 5 (5q12-5q14).";  
 RL Genomics 14:845-851(1992).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM V3).  
 RC TISSUE=Brain;  
 RX MEDLINE=95181355; PubMed=7876137;  
 RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;  
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M  
 RT without a chondroitin sulfate attachment in region in mouse and human  
 RT tissues.";  
 RL J. Biol. Chem. 270:3914-3918(1995).  
 RN [7]

RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).  
RC TISSUE-Aortic smooth muscle;  
RX MEDLINE=99327053; PubMed=10397680;  
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,  
R Wright T.N.;  
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";  
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).  
RN [8]  
RP PARTIAL SEQUENCE.  
RC TISSUE=Brain;  
RX MEDLINE=89174663; PubMed=2456833;  
RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;  
RT "Isolation and partial characterization of a glial  
hyaluronate-binding protein.";  
RL J. Biol. Chem. 264:5981-5987(1989).  
RN [9]  
RP TISSUE SPECIFICITY OF ISOFORMS.  
RX MEDLINE=96213482; PubMed=8627343;  
RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;  
RT "Differential expression of versican isoforms in brain tumors.";  
RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).  
CC -|- FUNCTION: May play a role in intercellular signaling and in  
connecting cells with the extracellular matrix. May take part in  
the regulation of cell motility, growth and differentiation. Binds  
hyaluronic acid.  
CC -|- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
CC -|- ALTERNATIVE PRODUCTS: At least 5 isoforms; V0 (shown here), V1,  
V2, V3 and Vint; are produced by alternative splicing.  
CC -|- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed  
in normal brain, gliomas, medulloblastomas, schwannomas,  
neurofibromas, and meningiomas; V2 is restricted to normal brain  
and gliomas; V3 is found in all these tissues except  
medulloblastomas.  
CC -|- DEVELOPMENTAL STAGE: Disappears after the cartilage development.  
CC -|- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -|- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
CC -|- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -|- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.  
CC -|- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U16306; AAA65018.1; -;  
CC EMBL; X15998; CAA34128.1; -;  
CC EMBL; S52488; AAB24878.1; -;  
CC EMBL; U26555; AAA67565.1; -;  
CC EMBL; D32039; BAA06801.1; -;  
CC EMBL; J02814; AAA36437.1; -;  
CC EMBL; AF084545; AAD48545.1; -;  
CC PIR; S06014; S06014.  
CC PIR; A29348; A29348.  
CC PIR; A30358; A30358.  
CC HSP; P01132; IEGF.  
CC Genew; HGNC:2464; CSPG2.  
CC MIM; 118661; -;  
CC InterPro; IPR000152; Asx\_hydroxyl.  
CC InterPro; IPR000561; EGF-like.  
CC InterPro; IPR000742; EGF\_2.  
CC InterPro; IPR001881; EGF\_Ca.  
CC InterPro; IPR001438; EGF\_II.  
CC InterPro; IPR003599; Ig.  
CC InterPro; IPR003006; Ig\_MHC.  
CC InterPro; IPR001304; Lectin\_C.  
CC InterPro; IPR000538; Link.  
CC InterPro; IPR000436; Sushi\_SCR\_CCP.  
CC Pfam; PF00008; EGF; 2.

DR Pfam; PF00047; ig; 1.  
DR Pfam; PF00059; lectin\_C; 1.  
DR Pfam; PF00084; sushi; 1.  
DR Pfam; PF00193; Xlink; 2.  
DR PRINTS; PR00010; EGFBL00D.  
DR PRODOM; PD000918; Link; 2.  
DR SMART; SM00032; CCP; 1.  
DR SMART; SM00034; CLECT; 1.  
DR SMART; SM00179; EGF\_Ca; 1.  
DR SMART; SM00001; EGF\_like; 1.  
DR SMART; SM00409; IG; 1.  
DR SMART; SM00445; LINK; 2.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_Ca; 1.  
DR PROSITE; PS01241; LINK; 2.  
DR PROSITE; PS00615; C\_TYPE\_LECTIN\_1; 1.  
DR PROSITE; PS00041; C\_TYPE\_LECTIN\_2; 1.  
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;  
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
KW Hyaluronic acid; Alternative splicing.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 3396 VERSICAN CORE PROTEIN.  
FT DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 167 244 LINK 1.  
FT DOMAIN 265 346 LINK 2.  
FT DOMAIN 348 1335 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT  
DOMAIN).  
FT DOMAIN 1336 3089 GAG-BETA.  
FT DOMAIN 3089 3125 EGF-LIKE 1.  
FT DOMAIN 3127 3163 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 3176 3290 C-TYPE LECTIN.  
FT DOMAIN 3295 3353 SUSHI.  
FT DISULFID 44 130 BY SIMILARITY.  
FT DISULFID 172 243 BY SIMILARITY.  
FT DISULFID 196 217 BY SIMILARITY.  
FT DISULFID 270 345 BY SIMILARITY.  
FT DISULFID 294 315 BY SIMILARITY.  
FT DISULFID 3093 3104 BY SIMILARITY.  
FT DISULFID 3098 3113 BY SIMILARITY.  
FT DISULFID 3115 3124 BY SIMILARITY.  
FT DISULFID 3131 3142 BY SIMILARITY.  
FT DISULFID 3136 3151 BY SIMILARITY.  
FT DISULFID 3153 3162 BY SIMILARITY.  
FT DISULFID 3169 3180 BY SIMILARITY.  
FT DISULFID 3197 3289 BY SIMILARITY.  
FT DISULFID 3265 3281 BY SIMILARITY.  
FT DISULFID 3296 3339 BY SIMILARITY.  
FT DISULFID 3325 3352 BY SIMILARITY.  
FT CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 615 615 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 782 782 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 809 809 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1398 1398 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1442 1442 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1468 1468 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1663 1663 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1898 1898 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2179 2179 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2272 2272 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2280 2280 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2360 2360 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2385 2385 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2385 2385 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 86.0%;

Score 43; DB 1; Length 3396;

Best Local Similarity 70.0%;

Pred. No. 19;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

|||:|:|:|

Db 3115 CVPGYSGDQC 3124

RESULT 9  
FBP1\_STRPU STANDARD; PRT; 1064 AA.  
AC P10079;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Fibropellin I precursor (Epidermal growth factor-related protein 1)  
DE (UEGF-1).  
GN EGF1.  
OS Strongylocentrotus purpuratus (Purple sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;  
OC Strongylocentrotus.  
OX NCBI\_TaxID=7668;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90112459; PubMed=2514273;  
RA Delgadillo-Reynoso M.G., Rollo D.R., Hursh D.A., Raff R.A.;  
RT "Structural analysis of the uegf gene in the sea urchin  
RT strongylocentrotus purpuratus reveals more similarity to vertebrate  
RT than to invertebrate genes with EGF-like repeats.";  
RL J. Mol. Evol. 29:314-327(1989).  
RN [2]  
RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.  
RX MEDLINE=87319677; PubMed=3498216;  
RA Hursh D.A., Andrews M.E., Raff R.A.;  
RT "A sea urchin gene encodes a polypeptide homologous to epidermal  
RT growth factor.";  
RL Science 237:1487-1490(1987).  
RN [3]  
RP AVIDIN-LIKE DOMAIN  
RX MEDLINE=89196806; PubMed=2784773;  
RA Hunt L.T., Barker W.C.;  
RT "Avidin-like domain in an epidermal growth factor homolog from a sea  
RT urchin.";  
RL FASEB J. 3:1760-1764(1989).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=91285254; PubMed=2060714;  
RA Bisgrove B.W., Andrews M.E., Raff R.A.;  
RT "Fibropellins, products of an EGF repeat-containing gene, form a  
RT unique extracellular matrix structure that surrounds the sea urchin  
RT embryo.";  
RL Dev. Biol. 146:89-99(1991).  
CC -1- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR  
CC MATRIX.  
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM  
CC OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER  
CC THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE  
CC EMBRYOS AND EARLY LARVAE.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1A (SHOWN HERE) AND 1B; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (1B) LACKS 8 EGF  
CC REPEATS.  
CC -1- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND  
CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN  
CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS  
CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY  
CC AND ZYGOTICALLY.  
CC -1- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
CC -1- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR  
CC TO AVIDIN/STREPTAVIDIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL; L08692; AAA62164.1; -  
CC EMBL; L08692; AAA62163.1; -  
DR EMBL; X17530; CAA35571.1; -  
DR EMBL; M17421; AAA30050.1; -  
DR EMBL; X17533; CAA35573.1; -  
DR PIR; A29316; A29316.  
DR HSSP; P01132; LEGF.  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000088; Avidin.  
DR InterPro; IPR000859; CUB\_domain.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR000742; EGF\_2.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR001438; EGF\_II.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR Pfam; PF00008; EGF; 21.  
DR Pfam; PF00431; CUB; 1.  
DR Pfam; PF01382; Avidin; 1.  
DR PRINTS; PR00010; EGFBL00D.  
DR PRINTS; PR00011; EGF\_LAMININ.  
DR SMART; SM00042; CUB; 1.  
DR SMART; SM00179; EGF\_CA; 18.  
DR SMART; SM00001; EGF\_like; 3.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 19.  
DR PROSITE; PS00022; EGF\_1; 19.  
DR PROSITE; PS00577; AVIDIN; 1.  
DR PROSITE; PS01180; CUB; 1.  
DR PROSITE; PS01186; EGF\_2; 19.  
DR PROSITE; PS01187; EGF\_CA; 18.  
KW Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;  
KW Glycoprotein; Calcium-binding.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1064 FIBROPELLIN I.  
FT DOMAIN 20 55 EGF-LIKE 1.  
FT DOMAIN 62 175 CUB.  
FT DOMAIN 176 212 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 366 402 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 442 478 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 480 516 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 518 554 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 556 592 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 594 630 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 632 668 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 670 706 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 708 744 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 746 782 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 784 820 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 822 858 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 860 896 EGF-LIKE 20.  
FT DOMAIN 898 934 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 936 1064 AVIDIN-LIKE.  
FT DISULFID 23 34 BY SIMILARITY.  
FT DISULFID 28 43 BY SIMILARITY.  
FT DISULFID 45 54 BY SIMILARITY.  
FT DISULFID 180 191 BY SIMILARITY.  
FT DISULFID 185 200 BY SIMILARITY.  
FT DISULFID 202 211 BY SIMILARITY.  
FT DISULFID 218 229 BY SIMILARITY.  
FT DISULFID 223 238 BY SIMILARITY.  
FT DISULFID 240 249 BY SIMILARITY.  
FT DISULFID 256 267 BY SIMILARITY.  
FT DISULFID 261 276 BY SIMILARITY.  
FT DISULFID 278 287 BY SIMILARITY.  
FT DISULFID 294 305 BY SIMILARITY.  
FT DISULFID 299 314 BY SIMILARITY.  
FT DISULFID 316 325 BY SIMILARITY.



```

FT DISULFID 332 343 BY SIMILARITY.
FT DISULFID 337 352 BY SIMILARITY.
FT DISULFID 354 363 BY SIMILARITY.
FT DISULFID 370 381 BY SIMILARITY.
FT DISULFID 375 390 BY SIMILARITY.
FT DISULFID 392 401 BY SIMILARITY.
FT DISULFID 408 419 BY SIMILARITY.
FT DISULFID 413 428 BY SIMILARITY.
FT DISULFID 430 439 BY SIMILARITY.
FT DISULFID 446 457 BY SIMILARITY.
FT DISULFID 451 466 BY SIMILARITY.
FT DISULFID 468 477 BY SIMILARITY.
FT DISULFID 484 495 BY SIMILARITY.
FT DISULFID 489 504 BY SIMILARITY.
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 522 533 BY SIMILARITY.
FT DISULFID 527 542 BY SIMILARITY.
FT DISULFID 544 553 BY SIMILARITY.
FT DISULFID 560 571 BY SIMILARITY.
FT DISULFID 565 580 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 598 609 BY SIMILARITY.
FT DISULFID 603 618 BY SIMILARITY.
FT DISULFID 620 629 BY SIMILARITY.
FT DISULFID 636 647 BY SIMILARITY.
FT DISULFID 641 656 BY SIMILARITY.
FT DISULFID 658 667 BY SIMILARITY.
FT DISULFID 674 685 BY SIMILARITY.
FT DISULFID 679 694 BY SIMILARITY.
FT DISULFID 696 705 BY SIMILARITY.
FT DISULFID 712 723 BY SIMILARITY.
FT DISULFID 717 732 BY SIMILARITY.
FT DISULFID 734 743 BY SIMILARITY.
FT DISULFID 750 761 BY SIMILARITY.
FT DISULFID 755 770 BY SIMILARITY.
FT DISULFID 772 781 BY SIMILARITY.
FT DISULFID 788 799 BY SIMILARITY.
FT DISULFID 793 808 BY SIMILARITY.
FT DISULFID 810 819 BY SIMILARITY.
FT DISULFID 826 837 BY SIMILARITY.
FT DISULFID 831 846 BY SIMILARITY.
FT DISULFID 848 857 BY SIMILARITY.
FT DISULFID 864 875 BY SIMILARITY.
FT DISULFID 869 884 BY SIMILARITY.
FT DISULFID 886 895 BY SIMILARITY.
FT DISULFID 902 913 BY SIMILARITY.
FT DISULFID 907 922 BY SIMILARITY.
FT DISULFID 924 933 BY SIMILARITY.
FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 477 780 MISSING (IN ISOFORM IB).
FT CONFLICT 279 279 L -> S (IN REF. 2).
SQ SEQUENCE 1064 AA; 112072 MW; 2E569CA012ED6D09 CRC64;

Query Match 84.0%; Score 42; DB 1; Length 1064;
Best Local Similarity 60.0%; Pred. No. 9.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXYC 10
DB 316 CPLGFGDNC 325

RESULT 10
CLR2_MOUSE
ID CLR2_MOUSE STANDARD; PRT; 2920 AA.
AC Q9R0M0; Q9Z2R4; Q99K36;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Flamingo 1)
(mFmil).

```

```

GN CELSR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99418630; PubMed-10490098;
RA Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RA Takeichi M., Uemura T.;
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
RT polarity under the control of frizzled.";
RL Cell 98:585-595(1999).
RN [2]
RP SEQUENCE OF 1913-2796 FROM N.A., AND TISSUE SPECIFICITY.
RX PubMed-10790539;
RA Formstone C.J., Barclay J., Rees M., Little P.F.R.;
RT "Chromosomal localization of Celsr2 and Celsr3 in the mouse; Celsr3 is
RT a candidate for the tippy (tip) lethal mutant on chromosome 9.";
RN [3]
RP Mamm. Genome 11:392-394(2000).
RN [4]
RP SEQUENCE OF 2014-2920 FROM N.A.
RX TISSUE-Breast tumor;
RA Strausberg R.;
RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP DEVELOPMENTAL STAGE.
RX PubMed-11850187;
RA Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;
RT "Developmental expression profiles of Celsr (Flamingo) genes in the
RT mouse.";
RL Mech. Dev. 112:157-160(2002).
CC -|- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
CC -|- DEVELOPMENTAL STAGE: Predominantly expressed in the developing
CC CNS, the emerging dorsal root ganglia and cranial ganglia. In the
CC CNS, expression is uniform along the rostrocaudal axis. During
CC gastrulation, it is expressed within the anterior neural ectoderm.
CC At E10, expression is strong in the ventricular zones (VZ) in all
CC sectors of the brain, and lower in the marginal zones (MZ). It is
CC strong in E12 and E15, expression is prominent in the brain. It is
CC predominant in VZ, lower in MZ, except in telecephalic MZ where it is
CC differentiating fields than in VZ, and lower in
CC hemispheres, and to a lesser extent in the tectum and cerebellum.
CC A weak expression is also observed in the fetal lungs, kidney and
CC epithelia. In the newborn and postnatal stages, expression remains
CC restricted to the VZ as well as in migrating and postmigratory
CC cells throughout the brain.
CC -|- TISSUE SPECIFICITY: Expressed in the CNS and in the eye.
CC -|- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -|- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.
CC -|- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC -|- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB028499; BAA84070.1; -
CC EMBL; AF031573; AAC68837.1; -
CC EMBL; BC005499; AAH05499.1; -
CC HSPSP; BC00740; LEDM
CC MGD; MGI:1858235; Celsr2.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR002126; Cadherin.

```

DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR000832; GPCR secretin.  
 DR InterPro: IPR001879; hormn\_receptor.  
 DR InterPro: IPR002049; Laminin\_EGF.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR InterPro: IPR000203; PKD\_cys\_rich.  
 DR Pfam: PF000002; 7tm2; 1.  
 DR Pfam: PF000028; cadherin; 9.  
 DR Pfam: PF000008; EGF; 5.  
 DR Pfam: PF01825; GPS; 1.  
 DR Pfam: PF02793; HRM; 1.  
 DR Pfam: PF00054; laminin\_G; 2.  
 DR PRINTS: PR00205; CADHERIN.  
 DR PRINTS: PR00011; EGF\_LAMININ.  
 DR PRINTS: PR00249; GPCRSECRETIN.  
 DR SMART: SM00112; CA; 9.  
 DR SMART: SM00180; EGF\_Lam; 1.  
 DR SMART: SM00001; EGF\_like; 6.  
 DR SMART: SM00303; GPS; 1.  
 DR SMART: SM00008; HormR; 1.  
 DR SMART: SM00282; LamG; 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE: PS00232; CADHERIN\_1; 6.  
 DR PROSITE: PS0268; CADHERIN\_2; 9.  
 DR PROSITE: PS00022; EGF\_1; 6.  
 DR PROSITE: PS01186; EGF\_2; 2.  
 DR PROSITE: PS00649; G-PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.  
 DR PROSITE: PS00650; G-PROTEIN\_RECEP\_F2\_2; FALSE\_NEG.  
 DR PROSITE: PS0227; G-PROTEIN\_RECEP\_F2\_3; 1.  
 DR PROSITE: PS0261; G-PROTEIN\_RECEP\_F2\_4; 1.  
 DR PROSITE: PS01248; LAMININ\_TYPE\_EGF; 1.  
 DR PROSITE: PS0221; GPS; 1.  
 DR PROSITE: PS00025; LAM\_G\_DOMAIN; 2.  
 DR G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;  
 KW Developmental protein; Hydroxylation; Signal.  
 FT SIGNAL 1 31  
 FT CHAIN 32 2920  
 FT FT  
 FT DOMAIN 32 2381  
 FT TRANSMEM 2382 2402  
 FT DOMAIN 2403 2414  
 FT TRANSMEM 2415 2434  
 FT DOMAIN 2435 2439  
 FT TRANSMEM 2440 2460  
 FT DOMAIN 2461 2481  
 FT TRANSMEM 2482 2502  
 FT DOMAIN 2503 2519  
 FT TRANSMEM 2520 2540  
 FT DOMAIN 2541 2564  
 FT TRANSMEM 2565 2585  
 FT DOMAIN 2586 2592  
 FT TRANSMEM 2593 2613  
 FT DOMAIN 2614 2920  
 FT DOMAIN 182 289  
 FT DOMAIN 290 399  
 FT DOMAIN 400 506  
 FT DOMAIN 507 611  
 FT DOMAIN 612 713  
 FT DOMAIN 714 816  
 FT DOMAIN 817 922  
 FT DOMAIN 923 1024  
 FT DOMAIN 1029 1147  
 FT DOMAIN 1229 1287  
 FT DOMAIN 1289 1319  
 FT DOMAIN 1329 1367  
 FT DOMAIN 1368 1572  
 FT DOMAIN 1575 1611  
 FT DOMAIN 1615 1792  
 FT DOMAIN 1788 1830  
 FT DOMAIN 1831 1868  
 FT DOMAIN 1884 1923

FT DOMAIN 1924 1956  
 FT DOMAIN 2317 2369  
 FT DOMAIN 2744 2749  
 FT DISULFID 1293 1304  
 FT DISULFID 1298 1313  
 FT DISULFID 1315 1318  
 FT DISULFID 1333 1344  
 FT DISULFID 1338 1354  
 FT DISULFID 1356 1366  
 FT DISULFID 1579 1590  
 FT DISULFID 1584 1599  
 FT DISULFID 1601 1610  
 FT DISULFID 1792 1803  
 FT DISULFID 1798 1818  
 FT DISULFID 1820 1829  
 FT DISULFID 1833 1844  
 FT DISULFID 1838 1856  
 FT DISULFID 1858 1867  
 FT DISULFID 1888 1900  
 FT DISULFID 1890 1907  
 FT DISULFID 1909 1922  
 FT DISULFID 1925 1937  
 FT DISULFID 1927 1944  
 FT DISULFID 1946 1955  
 FT MOD\_RES 1592 1592  
 FT CARBOHYD 486 486  
 FT CARBOHYD 558 558  
 FT CARBOHYD 702 702  
 FT CARBOHYD 1037 1037  
 FT CARBOHYD 1077 1077  
 FT CARBOHYD 1183 1183  
 FT CARBOHYD 1213 1213  
 FT CARBOHYD 1502 1502  
 FT CARBOHYD 1566 1566  
 FT CARBOHYD 1742 1742  
 FT CARBOHYD 1828 1828  
 FT CARBOHYD 1901 1901  
 FT CARBOHYD 2025 2025  
 FT CARBOHYD 2044 2044  
 FT CARBOHYD 2062 2062  
 FT CARBOHYD 2324 2324  
 FT CARBOHYD 2346 2346  
 FT CONFLICT 2199 2199  
 FT CONFLICT 2283 2283  
 FT CONFLICT 2535 2535  
 FT CONFLICT 2571 2571  
 FT CONFLICT 2639 2639  
 Query Match 84.08; Score 42; DB 1; Length 2920;  
 Best Local Similarity 60.08; Pred. No. 25;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CVIGXSGDXC 10  
 II::: III  
 Db 1820 CVLGYGDNC 1829  
 RESULT 11  
 7B2\_XENLA  
 ID 7B2\_XENLA STANDARD; PRT; 161 AA.  
 AC P18844;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Neuroendocrine protein 7B2 (Secretogranin V) (Fragment).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_Taxid=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89231705; PubMed=2714283;







FT DISULFID 338 BY SIMILARITY. 349  
 FT DISULFID 343 BY SIMILARITY. 358  
 FT DISULFID 360 BY SIMILARITY. 369  
 FT DISULFID 375 BY SIMILARITY. 386  
 FT DISULFID 380 BY SIMILARITY. 397  
 FT DISULFID 399 BY SIMILARITY. 408  
 FT DISULFID 415 BY SIMILARITY. 428  
 FT DISULFID 422 BY SIMILARITY. 437  
 FT DISULFID 439 BY SIMILARITY. 448  
 FT DISULFID 455 BY SIMILARITY. 466  
 FT DISULFID 460 BY SIMILARITY. 475  
 FT DISULFID 477 BY SIMILARITY. 486  
 FT DISULFID 493 BY SIMILARITY. 503  
 FT DISULFID 498 BY SIMILARITY. 512  
 FT DISULFID 514 BY SIMILARITY. 523  
 FT DISULFID 530 BY SIMILARITY. 541  
 FT DISULFID 535 BY SIMILARITY. 550  
 FT DISULFID 552 BY SIMILARITY. 561  
 FT DISULFID 568 BY SIMILARITY. 578  
 FT DISULFID 573 BY SIMILARITY. 587  
 FT DISULFID 589 BY SIMILARITY. 598  
 FT DISULFID 605 BY SIMILARITY. 616  
 FT DISULFID 610 BY SIMILARITY. 625  
 FT DISULFID 627 BY SIMILARITY. 636  
 FT DISULFID 643 BY SIMILARITY. 653  
 FT DISULFID 664 BY SIMILARITY. 662  
 FT DISULFID 673 BY SIMILARITY. 673  
 FT DISULFID 680 BY SIMILARITY. 691  
 FT DISULFID 685 BY SIMILARITY. 700  
 FT DISULFID 702 BY SIMILARITY. 711  
 FT DISULFID 718 BY SIMILARITY. 728  
 FT DISULFID 723 BY SIMILARITY. 737  
 FT DISULFID 739 BY SIMILARITY. 748  
 FT DISULFID 755 BY SIMILARITY. 766  
 FT DISULFID 760 BY SIMILARITY. 775  
 FT DISULFID 777 BY SIMILARITY. 786  
 FT DISULFID 793 BY SIMILARITY. 804  
 FT DISULFID 798 BY SIMILARITY. 813  
 FT DISULFID 815 BY SIMILARITY. 824  
 FT DISULFID 831 BY SIMILARITY. 842  
 FT DISULFID 836 BY SIMILARITY. 853  
 FT DISULFID 855 BY SIMILARITY. 864  
 FT DISULFID 871 BY SIMILARITY. 882  
 FT DISULFID 876 BY SIMILARITY. 891  
 FT DISULFID 893 BY SIMILARITY. 902  
 FT DISULFID 909 BY SIMILARITY. 920  
 FT DISULFID 914 BY SIMILARITY. 929  
 FT DISULFID 931 BY SIMILARITY. 940  
 FT DISULFID 947 BY SIMILARITY. 958  
 FT DISULFID 952 BY SIMILARITY. 967  
 FT DISULFID 969 BY SIMILARITY. 978

Query Match 80.0%; Score 40; DB 1; Length 2437;  
 Best Local Similarity 60.0%; Pred. No. 47;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDC 10  
 |||:|:|:|  
 Db 322 CVNGWTGDC 331

RESULT 15  
 DL\_DROME STANDARD; PRT; 833 AA.  
 AC P10041; Q9VDY2; Q99108;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Neurogenic locus Delta protein precursor.  
 GN DL OR CG3619.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC NCBI\_TaxID=7227;  
 RN SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
 RC TISSUE=Embryo;  
 RA Vaessin H., Bremer K.A., Knust E., Campos-Ortega J.A.;  
 RT "The neurogenic gene Delta of Drosophila melanogaster is expressed in  
 neurogenic territories and encodes a putative transmembrane protein  
 with EGF-like repeats.";  
 RL EMBO J. 6:3431-3440(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R; TISSUE=Embryo;  
 RX MEDLINE=89196890; PubMed=3149249;  
 RA Kopczyński C.C., Alton A.K., Fechtel K., Kooh P.J., Muskavitch M.A.T.;  
 RT "Delta, a Drosophila neurogenic gene, is transcriptionally complex and  
 encodes a protein related to blood coagulation factors and epidermal  
 growth factor of vertebrates.";  
 RL Genes Dev. 2:1723-1735(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iqbalwam C.,  
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wasarman D.A., Weinstock G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [4]  
 RP SEQUENCE OF 422-621 FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=87218537; PubMed=3107986;  
 RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D., Vaessin H.,  
 Campos-Ortega J.A.;  
 RT "Egfr homologous sequences encoded in the genome of Drosophila  
 melanogaster, and their relation to neurogenic genes.";  
 RL EMBO J. 6:761-766(1987).  
 RN [5]  
 RP PATTERN OF TRANSCRIPTION, AND CHARACTERIZATION.  
 RX MEDLINE=91209246; PubMed=2128477;







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein : protein search, using sw model  
Run on: July 2, 2003, 07:25:01 ; Search time 41.3333 Seconds  
(without alignments)  
49.850 Million cell updates/sec

Title: US-09-673-785D-13  
Perfect score: 50  
Sequence: 1 CVIGXSGDXC 10

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL.21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	1200	11	Q8Vd07 mus musculu
2	45	90.0	2146	5	Q9VC97 drosophila
3	45	90.0	4006	11	O35452 mus musculu
4	45	90.0	4114	11	O54796 mus musculu
5	43	86.0	372	5	Q21756 caenorhabdi
6	43	86.0	1162	5	Q8WTP0 halocynthia
7	43	86.0	2531	5	O16004 lytechinus
8	42	84.0	191	5	Q8T521 caenorhabdi
9	42	84.0	193	5	Q8T520 caenorhabdi
10	42	84.0	193	5	Q8T522 caenorhabdi
11	42	84.0	193	5	Q8T523 caenorhabdi
12	42	84.0	615	13	O57409 brachydanio
13	42	84.0	963	5	Q9GPM9 caenorhabdi
14	42	84.0	963	5	Q9GPM9 caenorhabdi
15	42	84.0	1270	5	Q9GPN0 caenorhabdi
16	42	84.0	1531	11	Q9WVB5 mus musculu

# SUMMARIES

17	42	84.0	2920	11	Q9ROM0
18	41	82.0	2352	5	O61240
19	40	80.0	106	4	Q9UN93
20	40	80.0	308	4	Q9UN94
21	40	80.0	366	4	Q9UN95
22	40	80.0	601	10	Q9LYF1
23	40	80.0	794	5	Q8T4P0
24	40	80.0	877	4	Q9H3Q6
25	40	80.0	878	4	Q9GZ22
26	40	80.0	901	4	Q9H3Q7
27	40	80.0	957	4	Q9H195
28	40	80.0	1217	4	O14651
29	40	80.0	1530	11	Q9UKW9
30	40	80.0	1531	11	Q9WUG5
31	39	78.0	89	6	O28867
32	39	78.0	190	11	Q9JKW3
33	39	78.0	264	5	Q20043
34	39	78.0	265	11	Q9DCP5
35	39	78.0	278	11	Q9QXT5
36	39	78.0	315	2	O33537
37	39	78.0	379	5	Q19922
38	39	78.0	403	5	O18375
39	39	78.0	422	16	P72725
40	39	78.0	426	16	O8Z098
41	39	78.0	470	10	Q9SUL6
42	39	78.0	473	4	Q9BW29
43	39	78.0	473	4	Q12891
44	39	78.0	473	4	Q12891
45	39	78.0	559	5	Q9VZ44

# ALIGNMENTS

## RESULT 1

Q8VD07	Q8VD07	PRELIMINARY;	PRT; 1200 AA.
ID	Q8VD07		
AC	Q8VD07		
DT	01-MAR-2002 (T-EMBLrel. 20, Created)		
DT	01-MAR-2002 (T-EMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (T-EMBLrel. 21, Last annotation update)		
DE	Similar to epidermal growth factor.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=KIDNEY;		
RA	Straussberg R.;		
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC017681; AAH17681.1; -		
DR	InterPro; IPR000152; Asx_Hydroxyl.		
DR	InterPro; IPR000561; EGF-like.		
DR	InterPro; IPR001336; EGF 1.		
DR	InterPro; IPR001881; EGF-Ca.		
DR	InterPro; IPR000033; Ldl_receptor_rep.		
DR	Pfam; PF00008; EGF; 8.		
DR	Pfam; PF00058; ldl_recept_b; 7.		
DR	PRINTS; PR00009; EGF-TGF.		
DR	SMART; SM00181; EGF; 9.		
DR	SMART; SM00179; EGF-Ca; 8.		
DR	SMART; SM00135; LY; 9.		
DR	PROSITE; PS00010; ASX-HYDROXYL; UNKNOWN_3.		
DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.		
DR	PROSITE; PS01186; EGF_2; UNKNOWN_6.		
DR	PROSITE; PS01187; EGF-Ca; UNKNOWN_3.		
DR	SEQUENCE 1200 AA; 131317 MW; B6A44F7294746476 CRC64;		
SQ			

Query Match 100.0%; Score 50; DB 11; Length 1200;  
Best Local Similarity 80.0%; Pred. No. 0.91;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CVIGSGDXC 10
Db 992 CVIGSGDRC 1001

RESULT 2
Q3VC97 PRELIMINARY; PRT; 2146 AA.
ID Q9VC97
AC Q9VC97;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CRB protein.
GN CRB OR CG6383.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler K., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003747; AAF56276.1; -.
DR HSSP; P00740; IEDM.
DR FlyBase; FBgn0000368; crb.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR002049; Laminin_EGF.
DR Dr Pfam; PF0001791; Laminin_G.
DR Dr Pfam; PF000008; EGF; 26.
DR Dr Pfam; PF000054; laminin_G; 3.

```

---

```

DR PRINTS; PRO0010; EGFBL00D.
DR PRINTS; PRO0011; EGFLAMININ.
DR SMART; SM00179; EGF_CA; 11.
DR SMART; SM00001; EGF_like; 16.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 16.
DR PROSITE; PS00022; EGF_1; 25.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS01187; EGF_CA; 13.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 2146 AA; 233570 MW; 8E23B9E32B761115 CRC64;

Query Match 90.0%; Score 45; DB 5; Length 2146;
Best Local Similarity 60.0%; Pred. No. 15;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10
Db 569 CAVGSGDRC 578

RESULT 3
Q35452 PRELIMINARY; PRT; 4006 AA.
ID Q35452
AC Q35452;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Tenascin X.
GN TNX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
RA Loretz C., Schmidt S., Traicoff R., Zackrone K., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class III
region.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF030001; AAB82015.1; -.
DR HSSP; P02671; IFZD.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00008; EGF; 15.
DR Pfam; PF00147; fibrinogen_C; 1.
DR Pfam; PF00041; fn3; 30.
DR SMART; SM00181; EGF; 8.
DR SMART; SM00001; EGF_like; 2.
DR SMART; SM00186; FBG; 1.
DR SMART; SM00060; FN3; 24.
DR PROSITE; PS00022; EGF_1; UNKNOWN_19.
DR PROSITE; PS01186; EGF_2; 19.
DR PROSITE; PS00514; FIBRIN_AG_DOMAIN; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 4006 AA; 435471 MW; 553PBE873498A4FC CRC64;

Query Match 90.0%; Score 45; DB 11; Length 4006;
Best Local Similarity 60.0%; Pred. No. 28;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10
Db 532 CAVGSGDRC 541

RESULT 4
Q54796 PRELIMINARY; PRT; 4114 AA.
ID Q54796
AC Q54796;
DT 01-JUN-1998 (TREMBlrel. 06, Created)

```



```
RESULT 7
OL6004 ID O16004 PRELIMINARY; PRT; 2531 AA.
AC O16004;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Notch homolog.
OS Lytechinus variegatus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;
OX NCBI_TaxID=7654;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97454256; PubMed=9310331;
RX Sherwood D.R., McClay D.R.;
RT "Identification and localization of a sea urchin Notch homologue:
RT insights into vegetal plate regionalization and Notch receptor
RT regulation.";
RL Development 124:3363-3374(1997).
DR EMBL; AF000634; AAB82088.1; -.
DR HSP; P01132; 1EGF.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; ASX_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00066; notch; 3.
DR PRINTS; P01415; ANKYRIN.
DR PRINTS; P00010; EGF_BLOOD.
DR PRINTS; P01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00001; EGF_like; 11.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 10.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 21.
DR PROSITE; PS00022; EGF_1; UNKNOWN_33.
DR PROSITE; PS01186; EGF_2; 25.
DR PROSITE; PS01187; EGF_CA; 20.
KW ANK repeat; Calcium-binding; EGF-like domain; Glycoprotein;
KW Hydroxylation; Repeat.
SQ SEQUENCE 2531 AA; 273982 MW; 5BF42BEC627CA303 CRC64;

Query Match 86.0%; Score 43; DB 5; Length 2531;
Best Local Similarity 60.0%; Pred. No. 42;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDNC 10
DB 615 CPVGTSGDNC 624
I : : : : :

RESULT 8
Q8T521 ID Q8T521 PRELIMINARY; PRT; 191 AA.
AC Q8T521;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491463; AAM09702.1; -.
FT NON_TER 1
FT NON_TER 191
SQ SEQUENCE 191 AA; 20977 MW; 2EEB21B5FFA46470 CRC64;

Query Match 84.0%; Score 42; DB 5; Length 191;
Best Local Similarity 60.0%; Pred. No. 4.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDNC 10
DB 53 CPLGYSGDNC 62
I : : : : :

RESULT 9
Q8T520 ID Q8T520 PRELIMINARY; PRT; 191 AA.
AC Q8T520;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491463; AAM09703.1; -.
FT NON_TER 1
FT NON_TER 191
SQ SEQUENCE 191 AA; 20943 MW; 2EEB21B5F47FB470 CRC64;

Query Match 84.0%; Score 42; DB 5; Length 191;
Best Local Similarity 60.0%; Pred. No. 4.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDNC 10
DB 53 CPLGYSGDNC 62
I : : : : :

RESULT 10
Q8T522 ID Q8T522 PRELIMINARY; PRT; 193 AA.
AC Q8T522;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
```

```

DR EMBL; AF491460; AAM09700.1; -.
FT NON_TER 1
FT NON_TER 193
SQ SEQUENCE 193 AA; 21178 MW; 42135BB8E8BE02C5 CRC64;

Query Match      84.0%; Score 42; DB 5; Length 193;
Best Local Similarity 60.0%; Pred. No. 4.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
Db 55 CPLGYSGDYC 64

RESULT 11
Q8STG0 PRELIMINARY; PRT; 193 AA.
AC Q8STG0;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Glp-1 (Fragment)
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK104, AND HK105;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RL genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491459; AAM09699.1; -.
DR EMBL; AF491461; AAM09701.1; -.
FT NON_TER 1
FT NON_TER 193
SQ SEQUENCE 193 AA; 21180 MW; 413C8AB647B5C540 CRC64;

Query Match      84.0%; Score 42; DB 5; Length 193;
Best Local Similarity 60.0%; Pred. No. 4.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
Db 56 CPLGYSGDYC 65

RESULT 12
Q8T5Z3 PRELIMINARY; PRT; 194 AA.
AC Q8T5Z3;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Glp-1 (Fragment)
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PB826;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RL genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491458; AAM09698.1; -.
FT NON_TER 1
FT NON_TER 194
SQ SEQUENCE 194 AA; 21275 MW; 07313C0AB9C65C5 CRC64;

Query Match      84.0%; Score 42; DB 5; Length 194;

```

```

Best Local Similarity 60.0%; Pred. No. 4.8;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
Db 56 CPLGYSGDYC 65

RESULT 13
O57409 PRELIMINARY; PRT; 615 AA.
AC O57409;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE DeltaB.
DN DB OR DELTAB.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98165391; PubMed=9425132;
RA Haddon C., Smithers L., Schneider-Maunoury S., Coche T., Henrique D.,
RA Lewis J.;
RT "Multiple delta genes and lateral inhibition in zebrafish primary
RT neurogenesis.";
RL Development 125:359-370(1998).
DR EMBL; AF006488; AAC41241.1; -.
DR HSSP; P00740; 1EDM.
DR 2FIN; ZDB-GENE-980526-114; dlb.
DR InterPro; IPR00152; Asx_hydroxyl.
DR InterPro; IPR001774; DSL.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR001438; EGF_II.
DR Pfam; PF01414; DSL; 1.
DR Pfam; PF00008; EGF; 7.
DR PRINTS; PR00010; EGF_BLOOD.
DR SMART; SM00051; DSL; 1.
DR SMART; SM00179; EGF_CA; 3.
DR SMART; SM00001; EGF_like; 5.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 615 AA; 67592 MW; CA18004428F5603C CRC64;

Query Match      84.0%; Score 42; DB 13; Length 615;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
Db 194 CLVGWQGDYC 203

RESULT 14
Q9GPM9 PRELIMINARY; PRT; 963 AA.
AC Q9GPM9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Notch-like transmembrane receptor (Fragment).
DN GLP-1.
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.

```

```
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AF16;
RX MEDLINE=21100341; PubMed=11156985;
RA Rudel D., Kimble J.;
RT "Conservation of glp-1 regulation and function in nematodes.";
RL Genetics 157:639-654(2001).
DR EMBL; AF315555; AAG49317.1; -.
DR HSSP; P01132; IEGF.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00066; notch; 3.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00181; EGF; 5.
DR SMART; SM00001; EGF_like; 4.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 3.
DR ANK repeat; Receptor; Repeat; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 963 AA; 106394 MW; 52BF80010E12FB57 CRC64;

Query Match 84.0%; Score 42; DB 5; Length 963;
Best Local Similarity 60.0%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
Db 123 CPLGYSGDYC 132
|.:|:|:|

RESULT 15
QSGPNO
ID Q9GPN0 PRELIMINARY; PRT; 1270 AA.
AC Q9GPN0;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Notch-like transmembrane receptor.
GN GLP-1.
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AF16;
RX MEDLINE=21100341; PubMed=11156985;
RA Rudel D., Kimble J.;
RT "Conservation of glp-1 regulation and function in nematodes.";
RL Genetics 157:639-654(2001).
DR EMBL; AF315554; AAG49316.1; -.
DR HSSP; P01132; IEGF.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00023; ank; 6.
```

```
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00181; EGF; 12.
DR SMART; SM00179; EGF_CA; 10.
DR SMART; SM00001; EGF_like; 8.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_10.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 1.
DR ANK repeat; Receptor; Repeat; Transmembrane.
SQ SEQUENCE 1270 AA; 138964 MW; A7662EB575A4B61B CRC64;

Query Match 84.0%; Score 42; DB 5; Length 1270;
Best Local Similarity 60.0%; Pred. No. 32;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
Db 430 CPLGYSGDYC 439
|.:|:|:|

Search completed: July 2, 2003, 07:28:39
Job time : 42.3333 secs
```